JC639 U.S

A-REISSUE SQ Listing

PATENT

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re U.S. Letters Patent of ALBERTSEN et al. U.S. Letters Patent No. 5,691,454 (Serial No. 08/452,654)) ATTN: Applications Branch) Previous Examiner: N. Johnson)	JC542 U.S. PTO 09/442489
Issued: November 25, 1997 (Filed: May 25, 1995)))) Atty. Dkt. No. 01107.78817	
For: APC ANTIBODIES		

SUBMISSION OF REISSUE APPLICATION

Assistant Commissioner of Patents Washington, D.C. 20231

Sir:

A reissue application is hereby requested on behalf of the current assignees of record, Zeneca, Ltd.; The Cancer Institute, Japanese Foundation for Cancer Research; The Johns Hopkins University; and the University of Utah. Accompanying this submission are:

- a reissue application under 37 C.F.R. § 1.173;
- an amendment under 37 C.F.R. § 1.121(b);
- a computer readable form and paper copy of a substitute sequence listing;
- a reissue declaration; and
- assent of all assignees of record.

Transfer of all formal drawings from the patent file is requested. Copies of the formal drawings are enclosed for the Examiner's convenience.

Assignees offer to surrender the original patent upon indication of allowance of this application.

Respectfully submitted,

Date: November 18, 1999

Lisa M. Hemmendinger Registration No. 42,653

M. Henmendiscen

Banner & Witcoff, Ltd. 1001 G Street, N.W., Eleventh Floor Washington, D.C. 20001-4597 (202) 508-9100

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of)
ALBERTSON et al.) Group Art Unit: T.B.A.
ALBERTSON Ci ai.) Examiner: T.B.A.
Serial No. T.B.A.	j
Filed: even herewith)) Atty. Dkt. No. 01107.78817
For: APC ANTIBODIES	

AMENDMENT UNDER 37 C.F.R. § 1.121(b)

Assistant Commissioner of Patents Washington, D.C. 20231

Sir:

Please enter the following amendments to the resisue application referenced above. We believe no fee is due in connection with this amendment. If a fee is due, please charge Deposit Account No. 19-0733.

IN THE SPECIFICATION

At column 3, line 20:

In even another embodiment a preparation of the human APC protein is provided which is substantially free of other human proteins. The amino acid sequence of the protein is shown in [FIG. 3] FIGS. 3A-3Z (SEQ ID NOS: 7 and 2).

At column 4, line 26:

[FIGS. 3A-3F] <u>FIGS. 3A-3Z</u> show the sequence of the APC gene product (SEQ ID NO: 7). The cDNA sequence was determined through the analysis of 87 cDNA clones derived from normal colon, liver, and brain. A total of 8973 bp were contained within overlapping cDNA

clones, defining an ORF of [2842] <u>2843</u> amino acids. In frame stop codons surrounded this ORF, as described in the text, suggesting that the entire APC gene product was represented in the ORF illustrated. Only the predicted amino acids are shown.

At column 6, line 30:

Alteration of wild-type genes can also be detected on the basis of the alteration of a wild-type expression product of the gene. Such expression products include both the APC mRNA as well as the APC protein product. The sequences of these products are shown in [FIG. 3] FIGS.

3A-3Z. Point mutations may be detected by amplifying and sequencing the mRNA or via molecular cloning of cDNA made from the mRNA. The sequence of the cloned cDNA can be determined using DNA sequencing techniques which are well known in the art. The cDNA can also be sequenced via the polymerase chain reaction (PCR) which will be discussed in more detail below.

At column 8, line 32:

In order to facilitate subsequent cloning of amplified sequences, primers may have restriction enzyme site sequences appended to their 5' ends. Thus, all nucleotides of the primers are derived from APC sequences or sequences adjacent to APC except the few nucleotides necessary to form a restriction enzyme site. Such enzymes and sites are well known in the art. The primers themselves can be synthesized using techniques which are well known in the art. Generally, the primers can be made using oligonucleotide synthesizing machines which are commercially available. Given the sequence of the APC open reading frame shown in [FIG. 3] FIGS. 3A-3Z (SEQ ID NO: 1), design of particular primers is well within the skill of the art.

At column 10, line 39:

Polypeptides which have APC activity can be supplied to cells which carry mutant or missing APC alleles. The sequence of the APC protein is disclosed in [FIG. 3] FIGS. 3A-3Z (SEQ ID NO:7). [These two sequences differ slightly and appear to be indicate the existence of two different forms of the APC protein.] Protein can be produced by expression of the cDNA sequence in bacteria, for example, using known expression vectors. Alternatively, APC can be extracted from APC-producing mammalian cells such as brain cells. In addition, the techniques of synthetic chemistry can be employed to synthesize APC protein. Any of such techniques can provide the preparation of the present invention which comprises the APC protein. The preparation is substantially free of other human proteins. This is most readily accomplished by synthesis in a microorganism or in vitro.

At column 10, line 66:

A short region of homology has been identified between APC and the human m3 muscarinic acetylcholine receptor (mAChR). This homology was largely confined to 29 residues in which 6 out of 7 amino acids (EL(GorA)GLQA) were identical (See [FIG. 4] FIG. 4B (SEQ ID NO: 9)). Initially, it was not known whether this homology was significant, because many other proteins had higher levels of global homology (though few had six out of seven contiguous amino acids in common). However, a study on the sequence elements controlling G protein activation by mAChR subtypes (Lechleiter et al., EMBO J., p. 4381 (1990)) has shown that a 21 amino acid region from the m3 mAChR completely mediated G protein specificity when substituted for the 21 amino acids of m2 mAChR at the analogous protein position. These 21 residues overlap the 19 amino acid homology between APC and m3 mAChR.

At column 13, line 1:

Contig 2: TB1 - TB1 was identified through a cross-hybridization approach. Exons of genes are often evolutionarily conserved while introns and intergenic regions are much less conserved. Thus, if a human probe cross-hybridizes strongly to the DNA from non-primate species, there is a reasonable chance that it contains exon sequences. Subclones of the cosmids shown in [FIG. 1] FIGS. 1A, 1B-1, 1B-2, and 1B-3 were used to screen Southern blots containing rodent DNA samples. A subclone of cosmid N5.66 (p 5.66-4) was shown to strongly hybridize to rodent DNA, and this clone was used to screen cDNA libraries derived from normal adult colon and fetal liver. The ends of the initial cDNA clones obtained in this screen were then used to extend the cDNA sequence. Eventually, 11 cDNA clones were isolated, covering 2314 bp. The gene detected by these clones was named TB1. Sequence analysis of the overlapping clones revealed an open reading frame (ORF) that extended for 1302 bp starting from the most 5' sequence data obtained (FIG. 2A). If this entire open reading frame were translated, it would encode 434 amino acids (SEQ ID NO: 5). The product of this gene was not globally homologous to any other sequence in the current database but showed two significant local similarities to a family of ADP, ATP carrier/translocator proteins and mitochondrial brown fat uncoupling proteins which are widely distributed from yeast to mammals. These conserved regions of TB1 (underlined in FIG. 2A) may define a predictive motif for this sequence family. In addition, TB1 appeared to contain a signal peptide (or mitochondrial targeting sequence) as well as at least 7 transmembrane domains.

At column 14, line 38:

Sequence analysis of the APC cDNA clones revealed an open reading frame of 8,535 nucleotides. The 5' end of the ORF contained a methionine codon (codon 1) that was preceded

by an in-frame stop codon 9 bp upstream, and the 3' end was followed by several in-frame stop codons. The protein produced by initiation at codon 1 would contain [2,842] 2843 amino acids (SEQ ID NO: 7) [(FIG. 3)] FIG. 3A-3Z. The results of database searching with the APC gene product were quite complex due to the presence of large segments with locally biased amino acid compositions. In spite of this, APC could be roughly divided into two domains. The N-terminal 25% of the protein had a high content of leucine residues (12%) and showed local sequence similarities to myosins, various intermediate filament proteins (e.g., desmin, vimentin, neurofilaments) and Drosophila armadillo/human plakoglobin. The latter protein is a component of adhesive junctions (desmosomes) joining epithelial cells (Franke et al., Proc. Natl. Acad. Sci. U.S.A., Vol. 86, p. 4027 (1989); Perfer et al., Cell, Vol. 63, p. 1167 (1990)). The C-terminal 75% of APC (residues 731-2832) is 17% serine by composition with serine residues more or less uniformly distributed. This large domain also contains local concentrations of charged (mostly acidic) and proline residues. There was no indication of potential signal peptides, transmembrane regions, or nuclear targeting signals in APC, suggesting a cytoplasmic localization.

At column 26, line 27:

4 1 e

To obtain DNA sequence adjacent to the exons of the genes DP1, DP2.5, and SRP19, sequencing substrate was obtained by inverse PCR amplification of DNAs from two YACs, 310D8 and 183H12, that span the deletions. Ligation at low concentration cyclized the restriction enzyme-digested YAC DNAs. Oligonucleotides with sequencing tails, designed in inverse orientation at intervals along the cDNAs, primed PCR amplification from the cyclized templates. Comparison of these DNA sequences with the cDNA sequences placed exon boundaries at the divergence points. SRP19 and DP1 were each shown to have five exons. DP2.5 consisted of 15

exons. The sequences of the oligonucleotides synthesized to provide PCR amplification primers for the exons of each of these genes are listed in Table III [SEQ ID NOS:39-94] (SEQ ID NOS:39-94]. With the exception of exons 1, 3, 4, 9, and 15 of DP2.5 (see below), the primer sequences were located in intron sequences flanking the exons. The 5' primer of exon 1 is complementary to the cDNA sequence, but extends just into the 5' Kozak consensus sequence for the initiator methionine, allowing a survey of the translated sequences. The 5' primer of exon 3 is actually in the 5' coding sequences of this exon, as three separate intronic primers simply would not amplify. The 5' primer of exon 4 just overlaps the 5' end of this exon, and we thus fail to survey the 19 most 5' bases of this exon. For exon 9, two overlapping primer sets were used, such that each had one end within the exon. For exon 15, the large 3' exon of DP2.5, overlapping primer pairs were placed along the length of the exon; each pair amplified a product of 250-400 bases.

At column 29, line 1:

The sequences of the unique conformers from exons 7, 8, 10, and 11 of DP2.5 revealed dramatic mutations in the DP2.5 gene. The sequence of the new mutation creating the exon 7 conformer in patient 3746 was shown to contain a deletion of two adjacent nucleotides, at positions 730 and 731 in the cDNA sequence ([FIG. 7,] SEQ ID NO:1). The normal sequence at this splice junction is CAGGGTCA (intronic sequence underlined), with the intron-exon boundary between the two repetitions of AG. The mutant allele in this patient has the sequence CAGGTCA. Although this change is at the 5' splice site, comparison with known consensus sequences of splice junctions would suggest that a functional splice junction is maintained. If this new splice junction were functional, the mutation would introduce a frameshift that creates a stop

codon 15 nucleotides downstream. If the new splice junction were not functional, messenger processing would be significantly altered.

At column 29, line 26:

The unique conformer found in exon 8 of patient 3460 was found to carry a C-T transition, at position 904 in the cDNA sequence of DP2.5 [(shown in FIG. 7)], which replaced the normal sequence of CGA with TGA. This point mutation, when read in frame, results in a stop codon replacing the normal arginine codon. This single-base change had occurred within the context of a CG dimer, a potential hot spot for mutation (Barker et al., 1984).

At column 30, line 37:

The continuity of the very large (6.5 kb), most 3' exon in DP2.5 was shown in two ways. First, inverse PCR with primers spanning the entire length of this exon revealed no divergence of the cDNA sequence from the genomic sequence. Second, PCR amplification with converging primers placed at intervals along the exon generated products of the same size whether amplified from the originally isolated cDNA, cDNA from various tissues, or genomic template. Two forms of exon 9 were found in DP2.5: one is the complete exon; and the other, labeled exon 9A, is the result of a splice into the interior of the exon that deletes bases 934 to 1236 in the mRNA and removes 101 amino acids from the predicted protein (see [FIG. 3] <u>FIGS. 3A-3Z</u>, SEQ ID NOS: 1 & 2).

At column 31, line 30:

The cDNA consensus sequence of APC predicts that the longer, more abundant form of the message codes for a [2842 or 2844] 2843 amino acid peptide with a mass of 311.8 kd. This predicted APC peptide was compared with the current data bases of protein and DNA sequences using both Intelligenetics and GCG software packages. No genes with a high degree of amino

acid sequence similarity were found. Although many short (approximately 20 amino acid) regions of sequence similarity were uncovered, none was sufficiently strong to reveal which, if any, might represent functional homology. Interestingly, multiple similarities to myosins and keratins did appear. The APC gene also was scanned for sequence motifs of known function; although multiple glycosylation, phosphorylation, and myristoylation sites were seen, their significance is uncertain.

At columns 31-132:

Please delete the sequence listing and replace it with the enclosed substitute sequence listing. The substitute sequence listing is identical to the sequence listing in the patent with the exception of one amino acid in SEQ ID NO:7. The substitute sequence listing contains a proline at position 173.

Remarks

The specification has been amended to correct the number of amino acids said to be present in the APC protein. This correction is supported in Figure 3 and in SEQ ID NOS:1 and 2, each of which show a 2843 amino acid APC protein.

The sequence listing has been amended to correct the amino acid sequence of the APC protein shown in SEQ ID NO:7, by insertion of a proline at position 173 of SEQ ID NO:7. This amendment is supported in the issued patent in Figure 3 and in SEQ ID NOS:1 and 2, each of which contain a proline at position 173. A computer readable form of the substitute sequence listing is provided for use in examining this application. The contents of the computer readable form and the paper copy of the substitute sequence listing are identical. The contents of the substitute sequence listing are identical to those of the original sequence listing except for the insertion of the proline at position 173

in SEQ ID NO:7.

The specification also has been amended to refer separately to each figure according to 37 C.F.R. § 1.74 and to delete references to originally filed Figure 7, which was cancelled during prosecution.

None of the amendments to the specification or sequence listing adds new matter.

Respectfully submitted,

Date: November 18, 1999

Lisa M. Hemmendinger
Registration No. 42,653

Banner & Witcoff, Ltd. 1001 G Street, N.W., Eleventh Floor Washington, D.C. 20001-4597 (202) 508-9100

1 APC ANTIBODIES

This application is a division, of application Ser. No. 08/289,548, filed Aug. 12, 1994, which is a division of application Ser. No. 07/741,940 filed Aug. 8, 1991 (issued as 5 U.S. Pat. No. 5,352,775).

The U.S. Government has a paid-up license in this invention and the right in limited circumstances to require the patent owner to license others on reasonable terms as provided for by the terms of grants awarded by the National 10 Institutes or Health.

TECHNICAL AREA OF THE INVENTION

The invention relates to the area of cancer diagnostics and therapeutics. More particularly, the invention relates to detection of the germline and somatic alterations of wild-type APC genes. In addition, it relates to therapeutic intervention to restore the function of APC gene product.

BACKGROUND OF THE INVENTION 20

According to the model of Knudson for tumorigenesis (Cancer Research, Vol. 45, p. 1482, 1985), there are tumor suppressor genes in all normal cells which, when they become non-functional due to mutation, cause neoplastic development. Evidence for this model has been found in the cases of retinoblastoma and colorectal tumors. The implicated suppressor genes in those tumors, RB, p53, DCC and MCC, were found to be deleted or altered in many cases of the tumors studied. (Hansen and Cavenee, Cancer Research, Vol. 47, pp: 5518-5527 (1987); Baker et al., Science, Vol. 244, p. 217 (1989); Fearon et al., Science, Vol. 247, p. 49 (1990); Kinzler et al. Science Vol. 251. p. 1366 (1991).)

In order to fully understand the pathogenesis of tumors, it will be necessary to identify the other suppressor genes that 35 play a role in the tumorigenesis process. Prominent among these is the one(s) presumptively located at 5q21. Cytogenetic (Herrera et al., Am J. Med. Genet., Vol. 25, p. 473 (1986) and linkage (Leppert et al., Science, Vol. 238, p. 1411 (1987); Bodmer et al., Nature, Vol. 328, p. 614 (1987)) 40 studies have shown that this chromosome region harbors the gene responsible for familial adenomatous polyposis (FAP) and Gardner's Syndrome (GS). FAP is an autosomaldominant, inherited disease in which affected individuals develop hundreds to thousands of adenomatous polyps, 45 some of which progress to malignancy. GS is a variant of FAP in which desmold tumors, osteomas and other soft tissue tumors occur together with multiple adenomas of the colon and rectum. A less severe form of polyposis has been identified in which only a few (2-40) polyps develop. This 50 condition also is familial and is linked to the same chromosomal markers as FAP and GS (Leppert et al., New England Journal of Medicine, Vol. 322, pp. 904-908, 1990.) Additionally, this chromosomal region is often deleted from the adenomas (Vogelstein et al., N. Engl. J. Med., Vol. 319, 55 p. 525 (1988)) and carcinomas (Vogelstein et al., N. Engl. J. Med., Vol. 319, p. 525 (1988); Solomon et al., Nature, Vol. 328, p. 616 (1987); Sasaki et al., Cancer Research, Vol. 49, p. 4402 (1989); Delattre et al., Lancet, Vol. 2, p. 353 (1989); and Ashton-Rickardt et al., Oncogene, Vol. 4, p. 1169 60 (1989)) of patients without FAP (sporadic tumors). Thus, a putative suppressor gene on chromosome 5q21 appears to play a role in the early stages of colorectal neoplasia in beth sporadic and familial tumors.

Although the MCC gene has been identified on 5q21 as a 65 candidate suppressor gene, it does not appear to be altered in FAP or GS patients. Thus there is a need in the art for

investigations of this chromosomal region to identify genes and to determine if any of such genes are associated with FAP and/or GS and the process of tumorigenesis.

SUMMARY OF THE INVENTION

5

It is an object of the present invention to provide a method for diagnosing and prognosing a neoplastic tissue of a human

It is another object of the invention to provide a method of detecting genetic predisposition to cancer.

It is another object of the invention to provide a method of supplying wild-type APC gene function to a cell which has lest said gene function.

It is yet another object of the invention to provide a kit for determination of the nucleotide sequence of APC alleles by the polymerase chain reaction.

It is still another object of the invention to provide nucleic acid probes for detection of mutations in the human APC 20 gene.

It is still another object of the invention to provide a cDNA molecule encoding the APC gene product.

It is yet another object of the invention to provide a preparation of the human APC protein.

It is another object of the invention to provide a method of screening for genetic prodisposition to cancer.

It is an object of the invention to provide methods of testing therapeutic agents for the ability to suppress neopla-30 sia.

It is still another object of the invention to provide animals carrying mutant APC alleles.

These and other objects of the invention are provided by one or more of the embodiments which are described below. In one embodiment of the present invention a method of diagnosing or prognosing a neoplastic tissue of a human is provided comprising: detecting somatic alteration of wild-type APC genes or their expression products in a sporadic colorectal cancer tissue, said alteration indicating neoplasia of the tissue.

In yet another embodiment a method is provided of detecting genetic predisposition to cancer in a human including familial adenomatous polyposis (FAP) and Gardner's Syndrome (GS), comprising: isolating a human sample selected from the group consisting of blood and fetal tissue; detecting alteration of wild-type APC gene coding sequences or their expression products from the sample, said alteration indicating genetic predisposition to cancer.

In another embodiment of the present invention a method is provided for supplying wild-type APC gene function to a cell which has lost said gene function by virtue of a mutation in the APC gene, comprising: introducing a wild-type APC gene into a cell which has lost said gene function such that said wild-type gene is expressed in the cell.

In another embodiment a method of supplying wild-type APC gene function to a cell is provided comprising: introducing a portion of a wild-type APC gene into a cell which has lost said gene function such that said portion is expressed in the cell, said portion encoding a part of the APC protein which is required for non-neoplastic growth of said cell. APC protein can also be applied to cells or administered to animals to remediate for mutant APC genes. Synthetic peptides or drugs can also be used to mimic APC function in cells which have altered APC expression.

In yet another embodiment a pair of single stranded primers is provided for determination of the nucleotide sequence of the APC gene by polymerase chain reaction. The sequence of said pair of single stranded DNA primers is derived from chromosome 5q band 21, said pair of primers allowing synthesis of APC gene coding sequences.

In still another embodiment of the invention a nucleic acid probe is provided which is complementary to human wild-type APC gene ceding sequences and which can form mismatches with mutant APC genes, thereby allowing their detection by enzymatic or chemical cleavage or by shifts in electrophoretic mobility.

In another embodiment of the invention a method is provided for detecting the presence of a neoplastic tissue in a human. The method comprises isolating a body sample from a human; detecting in said sample alteration of a wild-type APC gene sequence or wild-type APC expression product, said alteration indicating the presence of a neoplastic tissue in the human.

In still another embodiment a cDNA molecule is provided which comprises the coding sequence of the APC gene.

In even another embodiment a preparation of the human APC protein is provided which is substantially free of other human proteins. The amino acid sequence of the protein is shown in FIG. 3 (SEQ ID NOS: 7 and 2).

In yet another embodiment of the invention a method is provided for screening for genetic predisposition to cancer, including familial adenomatous polyposis (FAP) and Gardner's Syndrome (GS), in a human. The method comprises: detecting among kindred persons the presence of a DNA polymorphism which is linked to a mutant APC allele in an individual having a genetic predisposition to cancer, said kindred being genetically related to the individual, the 30 presence of said polymorphism suggesting a predisposition to cancer.

In another embodiment of the invention a method of testing therapeutic agents for the ability to suppress a neoplastically transformed phenotype is provided. The 35 method comprises: applying a test substance to a cultured epithelial cell which carries a mutation in an APC allele; and determining whether said test substance suppresses the neoplastically transformed phenotype of the cell.

In another embodiment of the invention a method of testing therapeutic agents for the ability to suppress a neoplastically transformed phenotype is provided. The method comprises: administering a test substance to an animal which carries a mutant APC allele; and determining whether said test substance prevents or suppresses the growth of tumors.

In still other embodiments of the invention transgenic animals are provided. The animals carry a mutant APC allele from a second animal species or have been genetically engineered to contain an insertion mutation which disrupts an APC allele.

The present invention provides the art with the information that the APC gene, a heretofore unknown gene is, in fact, a target of mutational alterations on chromosome 5q21 and that these alterations are associated with the process of tumorigenesis. This information allows highly specific assays to be performed to assess the neoplastic status of a particular tissue or the predisposition to cancer of an individual. This invention has applicability to Familial Adenomatous Polyposis, sporadic colorectal cancers, Gardner's Syndrome, as well as the less severe familial polyposis discusses above.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1A shows an overview of yeast artificial chromosome (YAC) contigs. Genetic distances between selected 65 RFLP markers from within the contigs are shown in centi-Morgans.

FIGS. 1B-1, 1B-2 and 1B-3 show a detailed map of the three central contigs. The position of the six identified genes from within the FAP region is shown; the 5' and 3' ends of the transcripts from these genes have in general not yet been isolated, as indicated by the string of dots surrounding the bars denoting the genes' positions. Selected restriction endonuclease recognition sites are indicated. B. BssH2; S, SstII; M, MluI; N, NruI.

FIGS. 2A and 2B show the sequence of TB1 (FIG. 2A)
10 and TB2 (FIG. 2B) genes. The cDNA sequence of the TB1
gene was determined from the analysis of 11 cDNA clones
derived from normal colon and liver, as described in the text.
A total of 2314 bp were contained within the overlapping
cDNA clones, defining an ORF of 424 amino acids beginning at nucleotide 1. Only the predicted amino acids from
the ORF are shown. The carboxy-terminal end of the ORF
has apparently been identified, but the 5' end of the TB1
transcript has not yet been precisely determined.

The cDNA sequence of the TB2 gene was determined from the YS-39 clone derived as described in the text. This clone consisted of 2300 bp and defined an ORF of 185 amino acids beginning at nucleotide 1. Only the predicted amino acids are shown. The carboxy terminal end of the ORF has apparently been identified, but the 5' end of the TB2 transcript has not been precisely determined.

FIGS. 3A-3F show the sequence of the APC gene product (SEQ ID NO:7). The cDNA sequence was determined through the analysis of 87 cDNA clones derived from normal colon, liver, and brain. A total of 8973 bp were contained within overlapping cDNA clones, defining an ORF of 2842 amino acids. In frame stop codons surrounded this ORF, as described in the text, suggesting that the entire APC gene product was represented in the ORF illustrated. Only the predicted amino acids are shown.

FIGS. 4A and 4B show the local similarity between human APC (SEQ ID NO:2) and ral2 (SEQ ID NO:8) of yeast. FIG. 4A shows amino acids 203 to 233 of APC, and FIG. 4B shows amino acids 453 to 481 of APC. Local similarity among the APC (SEQ ID NO:2) and MCC genes (SEQ ID NO:10) genes and the m3 muscarinic acetylcholine receptor (SEQ ID NO:9) is shown. The region of the mAChR shown corresponds to that responsible for coupling the receptor to G proteins. The connecting lines indicate identities; dots indicate related amino acids residues.

FIG. 5 shows the genomic map of the 1200 kb NotI fragment at the FAP locus. The NotI fragment is shown as a bold line. Relevant parts of the deletion chromosomes from patients 3214 and 3824 are shown as stippled lines. Probes used to characterize the NotI fragment and the deletions, and three YACs from which subclones were obtained, are shown below the restriction map. The chimeric end of YAC 183H12 is indicated by a dotted line. The orientation and approximate position of MCC are indicated 55 above the map.

FIG. 6A-6D show the DNA sequence (SEQ ID NO:3) and predicted amino acid sequence of DP1 (TB2) (SEQ ID NO:4). The nucleotide numbering begins at the most 5' nucleotide isolated. A proposed initiation methionine (base 77) is indicated in bold type. The entire coding sequence is presented.

FIG. 7A, FIG. 7B-1, and FIG. 7B-2 show the arrangement of exons in DP2.5 (APC). (A) Exon 9 corresponds to nucleotides 933-1312; exon 9a corresponds to nucleotides 1236-1312. The stop codon in the cDNA is at nucleotide 8535. (B) Partial intronic sequence surrounding each exon is shown (SEQ ID NO: 11-38). 5' intron sequences of exons 2,

3. 4. 5. 6. 7. 8. 9, 10, 11, 12, 13, 14, and 15 are shown in SEQ ID NOS: 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, respectively. 3' intron sequences of exons 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, and 14 are shown in SEQ ID NOS: 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 5 respectively.

DETAILED DESCRIPTION

It is a discovery of the present invention that mutational events associated with tumorigenesis occur in a previously unknown gene on chromosome 5q named here the APC (Adenomatous Polyposis Coil) gene. Although it was previously known that deletion of alleles on chromosome 5q were common in certain types of cancers, it was not known that a target gene of these deletions was the APC gene. Further it was not known that other types of mutational events in the APC gene are also associated with cancers. The mutations of the APC gene can involve gross rearrangements, such as insertions and deletions. Point mutations have also been observed.

According to the diagnostic and prognostic method of the present invention, alteration of the wild-type APC gene is detected. "Alteration of a wild-type gene" according to the present invention encompasses all forms of mutations including deletions. The alteration may be due to either rearrangements such as insertions, inversions, and deletions, or to point mutations. Deletions may be of the entire gene or only a portion of the gene. Somatic mutations are those which occur only in certain tissues, e.g., in the tumor tissue, 30 and are not inherited in the germline. Germline mutations can be found in any of a body's tissues. If only a single allele is somatically mutated, an early neoplastic state is indicated. However, if both alleles are mutated then a late neoplastic state is indicated. The finding of APC mutations thus provides both diagnostic and prognostic information. An APC allele which is not deleted (e.g., that on the sister chromosome to a chromosome carrying an APC deletion) can be screened for other mutations, such as insertions, small deletions, and point mutations. It is believed that many 40 mutations found in tumor tissues will be those leading to decreased expression of the APC gene product. However, mutations leading to non-functional gene products would also lead to a cancerous state. Point mutational events may occur in regulatory regions, such as in the promoter of the 45 gene, leading to loss or diminution of expression of the mRNA...Point mutations may also abolish proper RNA processing, leading to loss of expression of the APC gene product.

In order to detect the alteration of the wild-type APC gene in a tissue, it is helpful to isolate the tissue free from surrounding normal tissues. Means for enriching a tissue preparation for tumor cells are known in the art. For example, the tissue may be isolated from paraffin or cryostat sections. Cancer cells may also be separated from normal cells by flow cytometry. These as well as other techniques for separating tumor from normal cells are well known in the art. If the tumor tissue is highly contaminated with normal cells, detection of mutations is more difficult.

Detection of point mutations may be accomplished by 60 molecular cloning of the APC allele (or alleles) and sequencing that allele(s) using techniques well known in the art. Alternatively, the polymerase chain reaction (PCR) can be used to amplify gene sequences directly from a genomic DNA preparation from the tumor tissue. The DNA sequence of the amplified sequences can then be determined. The polymerase chain reaction itself is well known in the art.

See, e.g., Saiki et al., Science, Vol. 239, p. 487, 1988; U.S. Pat. No. 4,683,203; and U.S. Pat. No. 4,683,195. Specific primers which can be used in order to amplify the gene will be discussed in more detail below. The ligase chain reaction. 5 which is known in the art, can also be used to amplify APC sequences. See Wu et al., Genomics, Vol. 4, pp. 560-569 (1989). In addition, a technique known as allele specific PCR can be used. (See Ruano and Kidd, Nucleic Acids Research, Vol. 17, p. 8392, 1989.) According to this 10 technique, primers are used which hybridize at their 3' ends to a particular APC mutation. If the particular APC mutation is not present, an amplification product is not observed. Amplification Refractory Mutation System (ARMS) can also be used as disclosed in European Patent Application 15 Publication No. 0332435 and in Newton et al., Nucleic Acids Research, Vol. 17, p.7, 1989. Insertions and deletions of genes can also be detected by cloning, sequencing and amplification. In addition, restriction fragment length polymorphism (RFLP) probes for the gene or surrounding 20 marker genes can be used to score alteration of an allele or an insertion in a polymorphic fragment. Such a method is particularly useful for screening among kindred persons of an affected individual for the presence of the APC mutation found in that individual. Single stranded conformation poly-25 morphism (SSCP) analysis can also be used to detect base change variants of an allele. (Orita et al., Proc. Natl. Acad. Sci. USA Vol. 86, pp. 2766–2770, 1989, and Genomics, Vol. 5, pp. 874-879, 1989.) Other techniques for detecting insertions and deletions as are known in the art can be used.

Alteration of wild-type genes can also be detected on the basis of the alteration of a wild-type expression product of the gene. Such expression products include both the APC mRNA as well as the APC protein product. The sequences of these products are shown in FIG. 3. Point mutations may be detected by amplifying and sequencing the mRNA or via molecular cloning of cDNA made from the mRNA. The sequence of the cloned cDNA can be determined using DNA sequencing techniques which are well known in the art. The cDNA can also be sequenced via the polymerase chain reaction (PCR) which will be discussed in more detail below.

Mismatches, according to the present invention are hybridized nucleic acid duplexes which are not 100% homologous. The lack of total homology may be due to 45 deletions, insertions, inversions, substitutions or frameshift mutations. Mismatch detection can be used to detect point mutations in the gene or its mRNA product. While these techniques are less sensitive than sequencing, they are simpler to perform on a large number of tumor samples. An 50 example of a mismatch cleavage technique is the RNase protection method, which is described in detail in Winter et al., Proc. Natl. Acad. Sci. USA, Vol. 82, p. 7575, 1985 and Meyers et al., Science, Vol. 230, p. 1242, 1985. In the practice of the present invention the method involves the use 55 of a labeled riboprobe which is complementary to the human wild-type APC gene coding sequence. The riboprobe and either mRNA or DNA isolated from the tumor tissue are annealed (hybridized) together and subsequently digested with the enzyme RNase A which is able to detect some 60 mismatches in a duplex RNA structure. If a mismatch is detected by RNase A, it cleaves at the site of the mismatch. Thus, when the annealed RNA preparation is separated on an electrophoretic gel matrix, if a mismatch has been detected and cleaved by RNase A, an RNA product will be seen 65 which is smaller than the full-length duplex RNA for the riboprobe and the mRNA or DNA. The riboprobe need not be the full length of the APC mRNA or gene but can be a

segment of either. II the riboprobe comprises only a segment of the APC mRNA or gene it will be desirable to use a number of these probes to screen the whole mRNA sequence for mismatches.

In similar fashion, DNA probes can be used to detect mismatches, through enzymatic or chemical cleavage. See, e.g., Cotton et al., Proc. Natl. Acad. Sci. USA, Vol. 85, 4397, 1988; and Shenk et al., Proc. Natl. Acad. Sci. USA, Vol. 72, p. 989, 1975. Alternatively, mismatches can be detected by shifts in the electrophoretic mobility of mismatched duplexes relative to matched duplexes. See, e.g., Cariello, Human Genetics, Vol. 42, p. 726, 1988. With either riboprobes or DNA probes, the cellular mRNA or DNA which might contain a mutation can be amplified using PCR (see below) before hybridization. Changes in DNA of the APC gene can also be detected using Southern hybridization, especially if the changes are gross rearrangements, such as deletions and insertions.

DNA sequences of the APC gene which have been amplified by use of polymerase chain reaction may also be screened using allele-specific probes. These probes are nucleic acid oligomers, each of which contains a region of the APC gene sequence harboring a known mutation. For example, one oligomer may be about 30 nucleotides in length, corresponding to a portion of the APC gene sequence. By use of a battery of such allele-specific probes, PCR amplification products can be screened to identify the presence of a previously identified mutation in the APC gene. Hybridization of allele-specific probes with amplified APC sequences can be performed, for example, on a nylon filter. Hybridization to a particular probe under stringent hybridization conditions indicates the presence of the same mutation in the tumor tissue as in the allele-specific probe.

Alteration of APC mRNA expression can be detected by any technique known in the art. These include Northern blot 35 analysis, PCR amplification and RNase protection. Diminished mRNA expression indicates an alteration of the wildtype APC gene. Alteration of wild-type APC genes can also be detected by screening for alteration of wild-type APC protein. For example, monoclonal antibodies immunoreac- 40 tive with APC can be used to screen a tissue. Lack of cognate antigen would indicate an APC mutation. Antibodies specific for products of mutant alleles could also be used to detect mutant APC gene product. Such immunological assays can be done in any convenient format known in the 45 art. These include Western blots, immunohistochemical assays and ELISA assays. Any means for detecting an altered APC protein can be used to detect alteration of wild-type APC genes. Functional assays can be used, such as protein binding determinations. For example, it is believed 50 that APC protein oligomerizes to itself and/or MCC protein or binds to a G protein. Thus, an assay for the ability to bind to wild type APC or MCC protein or that G protein can be employed. In addition, assays can be used which detect APC biochemical function. It is believed that APC is involved in 55 phospholipid metabolism. Thus, assaying the enzymatic products of the involved phospholipid metabolic pathway can be used to determine APC activity. Finding a mutant APC gene product indicates alteration of a wild-type APC

Mutant APC genes or gene products can also be detected in other human body samples, such as, scrum, stool, urine and sputum. The same techniques discussed above for detection of mutant APC genes or gene products in tissues can be applied to other body samples. Cancer cells are 65 sloughed off from tumors and appear in such body samples. In addition, the APC gene product itself may be secreted into

the extracellular space and found in these body samples even in the absence of cancer cells. By screening such body samples, a simple early diagnosis can be achieved for many types of cancers. In addition, the progress of chemotherapy or radiotherapy can be monitored more easily by testing such body samples for mutant APC genes or gene products.

The methods of diagnosis of the present invention are applicable to any tumor in which APC has a role in tumorigenesis. Deletions of chromosome arm 5q have been observed in tumors of lung, breast, colon, rectum, bladder, liver, sarcomas, stomach and prostate, as well as in leukemias and lymphomas. Thus these are likely to be tumors in which APC has a role. The diagnostic method of the present invention is useful for clinicians so that they can decide upon an appropriate course of treatment. For example, a tumor displaying alteration of both APC alleles might suggest a more aggressive therapeutic regimen than a tumor displaying alteration of only one APC allele.

The primer pairs of the present invention are useful for determination of the nucleotide sequence of a particular APC allele using the polymerase chain reaction. The pairs of single stranded DNA primers can be annealed to sequences within or surrounding the APC gene on chromosome 5q in order to prime amplifying DNA synthesis of the APC gene itself. A complete set of these primers allows synthesis of all of the nucleotides of the APC gene coding sequences, i.e., the exons. The set of primers preferably allows synthesis of both intron and exon sequences. Allele specific primers can also be used. Such primers anneal only to particular APC mutant alleles, and thus will only amplify a product in the presence of the mutant allele as a template.

In order to facilitate subsequent cloning of amplified sequences, primers may have restriction enzyme site sequences appended to their 5' ends. Thus, all nucleotides of the primers are derived from APC sequences or sequences adjacent to APC except the few nucleotides necessary to form a restriction enzyme site. Such enzymes and sites are well known in the art. The primers themselves can be synthesized using techniques which are well known in the art. Generally, the primers can be made using oligonucleotide synthesizing machines which are commercially available. Given the sequence of the APC open reading frame shown in FIG. 3 (SEQ ID NO:1), design of particular primers is well within the skill of the art.

The nucleic acid probes provided by the present invention are useful for a number of purposes. They can be used in Southern hybridization to genomic DNA and in the RNase protection method for detecting point mutations already discussed above. The probes can be used to detect PCR 50 amplification products. They may also be used to detect mismatches with the APC gene or mRNA using other techniques. Mismatches can be detected using either enzymes (e.g., S1 nuclease), chemicals (e.g., hydroxylamine or osmium tetroxide and piperidine), or changes in electro-55 phoretic mobility of mismatched hybrids as compared to totally matched hybrids. These techniques are known in the art. See, Cotton, supra, Shenk, supra, Myers, supra, Winter, supra, and Novack et al., Proc. Natl. Acad. Sci. USA, Vol. 83, p. 586, 1986. Generally, the probes are complementary 60 to APC gene coding sequences, although probes to certain introns are also contemplated. An entire battery of nucleic acid probes is used to compose a kit for detecting alteration of wild-type APC genes. The kit allows for hybridization to the entire APC gene. The probes may overlap with each 65 other or be contiguous.

If a riboprobe is used to detect mismatches with mRNA, it is complementary to the mRNA of the human wild-type

APC gene. The riboprobe thus is an anti-sense probe in that it does not code for the APC protein because it is of the opposite polarity to the sense strand. The riboprobe generally will be labeled with a radioactive, colorimetric, or fluorometric material, which can be accomplished by any means known in the art. If the riboprobe is used to detect mismatches with DNA it can be of either polarity, sense or anti-sense. Similarly, DNA probes also may be used to detect mismatches.

Nucleic acid probes may also be complementary to mutant alleles of the APC gene. These are useful to detect similar mutations in other patients on the basis of hybridization rather than mismatches. These are discussed above and referred to as allele-specific probes. As mentioned above, the A PC probes can also be used in Southern hybridizations to genomic DNA to detect gross chromosomal changes such as deletions and insertions. The probes can also be used to select cDNA clones of APC genes from tumor and normal tissues. In addition, the probes can be used to detect APC mRNA in tissues to determine if expression is diminished as a result of alteration of wild-type APC genes.

According to the present invention a method is also provided of supplying wild-type APC function to a cell which carries mutant APC alleles. Supplying such function should suppress neoplastic growth of the recipient cells. The 25 wild-type APC gene or a part of the gene may be introduced into the cell in a vector such that the gene remains extrachromosomal. In such a situation the gene will be expressed by the cell from the extrachromosomal location. If a gene portion is introduced and expressed in a cell carrying a 30 mutant APC allele, the gene portion should encode a part of the APC protein which is required for non-neoplastic growth of the cell. More preferred is the situation where the wildtype APC gene or a part of it is introduced into the mutant cell in such a way that it recombines with the endogenous 35 mutant APC gene present in the cell. Such recombination requires a double recombination event which results in the correction of the APC gene mutation. Vectors for introduction of genes beth for recombination and for extrachromosomal maintenance are known in the art and any suitable 40 vector may be used. Methods for introducing DNA into cells such as electroporation, calcium phosphate co-precipitation and viral transduction are known in the art and the choice of method is within the competence of the routineer. Cells transformed with the wild-type A PC gene can be used as 45 model systems to study cancer remission and drug treatments which promote such remission.

Similarly, cells and animals which carry a mutant APC allele can be used as model systems to study and test for substances which have potential as therapeutic agents. The 50 cells are typically cultured epithelial cells. These may be isolated from individuals with APC mutations, either somatic or germline. Alternatively, the cell line can be engineered to carry the mutation in the APC allele. After a test substance is applied to the cells, the neoplastically transformed pheno-type of the cell will be determined. Any trait of neoplastically transformed cells can be assessed, including anchorage-independent growth, tumorigenicity in nude mice, invasiveness of cells, and growth factor dependence. Assays for each of these traits are known in the art. 60

Animals for testing therapeutic agents can be selected after mutagenesis of whole animals or after treatment of germline cells or zygotes. Such treatments include insertion of mutant A PC alleles, usually from a second animal species, as well as insertion of disrupted homologous genes. 65 Alternatively, the endogenous APC gene(s) of the animals may be disrupted by insertion or deletion mutation. After test

substances have been administered to the animals, the growth of tumors must be assessed. If the test substance prevents or suppresses the growth of tumors, then the test substance is a candidate therapeutic agent for the treatment of FAP and/or sporadic cancers.

Polypeptides which have APC activity can be supplied to cells which carry mutant or missing APC alleles. The sequence of the APC protein is disclosed in FIG. 3 (SEQ ID NO:7). These two sequences differ slightly and appear to be indicate the existence of two different forms of the APC protein. Protein can be produced by expression of the cDNA sequence in bacteria, for example, using known expression vectors. Alternatively, APC can be extracted from APC-producing mammalian cells such as brain cells. In addition, the techniques of synthetic chemistry can be employed to synthesize APC protein. Any of such techniques can provide the preparation of the present invention which comprises the APC protein. The preparation is substantially free of other human proteins. This is most readily accomplished by synthesis in a microorganism or in vitro.

Active APC molecules can be introduced into cells by microinjection or by use of liposomes, for example. Alternatively, some such active molecules may be taken up by cells, actively or by diffusion. Extracellular application of APC gene product may be sufficient to affect tumor growth. Supply of molecules with APC activity should lead to a partial reversal of the neoplastic state. Other molecules with APC activity may also be used to effect such a reversal, for example peptides, drugs, or organic compounds.

The present invention also provides a preparation of antibodies immunoreactive with a human APC protein. The antibodies may be polyclonal or monoclonal and may be raised against native APC protein, APC fusion proteins, or mutant APC proteins. The antibodies should be immunoreactive with APC epitopes, preferably epitopes not present on other human proteins. In a preferred embodiment of the invention the antibodies will immunoprecipitate APC proteins from solution as well as react with APC protein on Western or immunoblots of polyacrylamide gels. In another preferred embodiment, the antibodies will detect APC proteins in paraffin or frozen tissue sections, using immunocytochemical techniques. Techniques for raising and purifying antibodies are well known in the art and any such techniques may be chosen to achieve the preparation of the invention.

Predisposition to cancers as in FAP and GS can be ascertained by testing any tissue of a human for mutations of the APC gene. For example, a person who has inherited a germline APC mutation would be prone to develop cancers.

This can be determined by testing DNA from any tissue of the person's body. Most simply, blood can be drawn and DNA extracted from the cells of the blood. In addition, prenatal diagnosis can be accomplished by testing fetal cells, placental cells, or amniotic fluid for mutations of the APC gene. Alteration of a wild-type APC allele, whether for example, by point mutation or by deletion, can be detected by any of the means discussed above.

Molecules of cDNA according to the present invention are intron-free, APC gene ceding molecules. They can be made by reverse transcriptase using the APC mRNA as a template. These molecules can be propagated in vectors and cell lines as is known in the art. Such molecules have the sequence shown in SEQ ID NO:3. The cDNA can also be made using the techniques of synthetic chemistry given the sequence disclosed herein.

A short region of homology has been identified between APC and the human m3 muscarinic acetylcholine receptor

(mAChR). This homology was largely confined to 29 residues in which 6 out of 7 amino acids (EL(GorA)GLQA) were identical (See FIG. 4 (SEQ ID NO: 9)). Initially, it was not known whether this homology was significant, because many other proteins had higher levels of global homology 5 (though few had six out of seven contiguous amino acids in common). However, a study on the sequence elements controlling G protein activation by mAChR subtypes (Lechleiter et al., EMBO J., p. 4381 (1990)) has shown that a 21 amino acid region from the m3 mAChR completely 10 mediated G protein specificity when substituted for the 21 amino acids of m2 mA ChR at the analogous protein position. These 21 residues overlap the 19 amino acid homology between APC and m3 mA ChR.

This connection between APC and the G protein activat- 15 ing region of mAChR is intriguing in light of previous investigations relating G proteins to cancer. For example, the RAS oncogenes, which are often mutated in colorectal cancers (Vogelstein, et al., N. Engl. J. Med., Vol. 319, p. 525 (1988); Bos et al., Nature Vol. 327, p. 293 (1987)), are 20 members of the (1 protein family (Bourne, et al, Nature, Vol. 348, p. 125 (1990)) as is an in vitro transformation suppressor (Noda et al., Proc. Natl. Acad. Sci. USA, Vol. 86, p. 162 (1989)) and genes mutated in hormone producing tumors (Candis et al., Nature, Vol. 340, p. 692 (1989); Lyons et al., 25 Science, Vol. 249, p. 655 (1990)). Additionally, the gene responsible for neurofibromatosis (presumably a tumor suppressor gene) has been shown to activate the GIPase activity of RAS (Xu et al., Cell, Vol. 63, p. 835 (1990); Martin et al., Cell, Vol. 63, p. 843 (1990); Ballester et al., 30 Cell, Vol. 63, p. 851 (1990)). Another interesting link between G proteins and colon cancer involves the drug sulindac. This agent has been shown to inhibit the growth of benign colon tumors in patients with FAP, presumably by virtue of its activity as a cyclooxygenase inhibitor (Waddell 35 et al., J. Surg. Oncology 24(1), 83 (1983); Wadell, et al., Am. J. Surg., 157(1), 175 (1989); Charneau et al., Gastroenterologie Clinique at Biologique 14(2), 153 (1990)). Cyclooxygenase is required to convert arachidonic acid to prostaglandins and other biologically active molecules. G proteins 40 are known to regulate phospholipase A2 activity, which generates arachidonic acid from phospholipids (Role et al., Proc. Natl. Acad. Sci. USA, Vol. 84, p. 3623 (1987); Kurachi et al., Nature, Vol. 337, 12 555 (1989)). Therefore we propose that wild-type APC protein functions by interacting 45 with a G protein and is involved in phospholipid metabo-

The following are provided for exemplification purposes only and are not intended to limit the scope of the invention which has been described in broad terms above.

EXAMPLE 1

This example demonstrates the isolation of a 5.5 Mb region of human DNA linked to the FAP locus. Six genes are 55 identified in this region, all of which are expressed in normal colon cells and in colorectal, lung, ad bladder tumors.

The cosmid markers YN5.64 and YN5.48 have previously been shown to delimit an 8 cM region containing the locus for FAP (Nakamura et al., Am. J. Hum. Genet. Vol. 43, p. 60 638 (1988)). Further linkage and pulse-field gel electrophoresis (PFGE) analysis with additional markers has shown that the FAP locus is contained within a 4 cM region bordered by cosmids EF5.44 and L5.99. In order to isolate clones representing a significant portion of this locus, a yeast 65 artificial chromosome (YAC) library was screened with various 5q21 markers. Twenty-one YAC clones, distributed

within six contigs and including 5.5 Mb from the region between YN5.64 and YN5.48, were obtained (FIG. 1A).

Three contigs encompassing approximately 4 Mb were contained within the central portion of this region. The YAC's constituting these contigs, together with the markers used for their isolation and orientations, are shown in FIG. 1. These YAC contigs were obtained in the following way. To initiate each contig, the sequence of a genomic marker cloned from chromosome 5q21 was determined and used to design primers for PCR. PCR was then carried out on pools of YAC clones distributed in microtiter trays as previously described (Anand et al., Nucleic Acids Research, Vol. 18, p. 1951 (1980)). Individual YAC clones from the positive pools were identified by further PCR or hybridization based assays, and the YAC sizes were determined by PFGE.

To extend the areas covered by the original YAC clones, "chromosomal walking" was performed. For this purpose, YAC termini were isolated by a PCR based method and sequenced (Riley et al., Nucleic Acids Research, Vol. 18, p. 2887 (1990)). PCR primers based on these sequences were then used to rescreen the YAC library. For example, the sequence from an intron of the FER gene (Hao et al., Mol. Cell. Biol., Vol. 9, p. 1587 (1989)) was used to design PCR primers for isolation of the 28EC1 and 5EH8 YACs. The termini of the 28EC1 YAC were sequenced to derive markers RHE28 and LHE28, respectively. The sequences of these two markers were then used to isolate YAC clones 15CH12 (from RHE28) and 40CF1 and 29EF1 (from LHE28). These five YAC's formed a contig encompassing 1200 kb (contig 1, FIG. 1B).

Similarly, contig 2 was initiated using cosmid N5.66 sequences, and contig 3 was initiated using sequences both from the MCC gene and from cosmid EF5.44. A walk in the telomeric direction from YAC 14FH1 and a walk in the opposite direction from YAC 39GG3 allowed connection of the initial contig 3 clones through YAC 37HG4 (FIG. 1B). YAC37HG4 was deposited at the National Collection of Industrial and Marine Bacteria (NCIMB), P.O. Box 31, 23 St. Machar Drive, Aberdeén AB2 1RY, Scotland, under Accession No. 40353 on Dec. 17, 1990.

Multipoint linkage analysis with the various markers used to define the contigs, combined with PFGE analysis, showed that contigs 1 and 2 were centromeric to contig 3. These contigs were used as tools to orient and/or identify genes which might be responsible for FAP. Six genes were found to lie within this cluster of YAC's, as follows:

Contig #1: FER-The FER gene was discovered through its homology to the viral oncogene ABL (Hao et al., supra). It has an intrinsic tyrosine kinase activity, and in situ 50 hybridization with an FER probe showed that the gene was located at 5q11-23 (Morris et al., Cytogenet. Cell. Genet., Vol. 53, p. 4, (1990)). Because of the potential role of this oncogene-related gene in neoplasia, we decided to evaluate it further with regards to the FAP locus. A human genomic 55 clone from FER was isolated (MF 2.3) and used to define a restriction fragment length polymorphism (RFLP), and the RFLP in turn used to map FER by linkage analysis using a panel of three generation families. This showed that FER was very tightly linked to previously defined polymorphic 60 markers for the FAP locus. The genetic mapping of FER was complemented by physical mapping using the YAC clones derived from FER sequences (FIG. 1B). Analysis of YAC contig 1 showed that FER was within 600 kb of cosmid marker M5.28, which maps to within 1.5 Mb of cosmid 65 L5.99 by PFGE of human genomic DNA. Thus, the YAC mapping results were consistent with the FER linkage data and PFGE analyses.

Contig 2: TB1-TB1 was identified through a crosshybridization approach. Exons of genes are often evolutionarily conserved while introns and intergenie regions are much less conserved. Thus, it a human probe crosshybridizes strongly to the DNA from non-primate species, 5 there is a reasonable chance that it contains exon sequences. Subclones of the cosmids shown in FIG. 1 were used to screen Southern blots containing rodent DNA samples. A subclone of cosmid N5.66 (p 5.66-4) was shown to strongly hvbridize to rodent DNA, and this clone was used to screen $_{10}$ cDNA libraries derived from normal adult colon and fetal liver. The ends of the initial cDNA clones obtained in this screen were then used to extend the cDNA sequence. Eventually, 11 cDNA clones were isolated, covering 2314 bp. The gene detected by these clones was named TB1. 15 Sequence analysis of the overlapping clones revealed an open reading frame (ORF) that extended for 1302 bp starting from the most 5' sequence data obtained (FIG. 2A). If this entire open reading frame were translated, it would encode 434 amino acids (SEQ ID NO:5). The product of this gene 20 was not globally homologous to any other sequence in the current database but showed two significant local similarities to a family of ADP, ATP carrier/translocator proteins and mitochondrial brown fat uncoupling proteins which are widely distributed from yeast to mammals. These conserved 25 regions of TB1 (underlined in FIG. 2A) may define a predictive motif for this sequence family. In addition, TB1 appeared to contain a signal peptide (or mitochondrial targeting sequence) as well as at least 7 transmembrane domains.

Contig 3: MCC, TB2, SRP and APC-The MCC gene was also discovered through a cross-hybridization approach, as described previously (Kinzler et al., Science Vol. 251, p. 1366 (1991)). The MCC gene was considered a candidate for causing FAP by virtue of its tight genetic linkage to FAP 35 susceptibility and its somatic mutation in sporadic colorectal carcinomas. However, mapping experiments suggested that the ceding region of MCC was approximately 50 kb proximal to the centromeric end of a 200 kb deletion found in an FAP patient. MCC cDNA probes detected a 10 kb mRNA 40 transcript on Northern blot analysis of which 4151 bp, including the entire open reading frame, have been cloned. Although the 3' non-translated portion or an alternatively spliced form of MCC might have extended into this deletion, it was possible that the deletion did not affect the MCC gene 45 product. We therefore used MCC sequences to initiate a YAC contig, and subsequently used the YAC clones to identify genes 50 to 250 kb distal to MCC that might be contained within the deletion.

In a first approach, the insert from YAC24ED6 (FIG. 1B) 50 was radiolabelled and hybridized to a cDNA library from normal colon. One of the cDNA clones (YS39) identified in this manner detected a 3.1 kb mRNA transcript when used as a probe for Northern blot hybridization. Sequence analysis of the YS39 clone revealed that it encompassed 2283 nucleotides and contained an ORF that extended for 555 bp from the most 5' sequence data obtained. If all of this ORF were translated, it would encode 185 amino acids (SEQ ID NO:6) (FIG. 2B). The gene detected by YS39 was named TB2. Searches of nucleotide and protein databases revealed 60 that the TB2 gene was not identical to any previously reported sequences nor were there any striking similarities.

Another clone (YS11) identified through the YAC 24ED6 screen appeared to contain portions of two distinct genes. Sequences from one end of YS11 were identical to at least 65 180 bp of the signal recognition particle protein SRP19 (Lingelbach et al. Nucleic Acids Research, Vol. 16, p. 9431

(1988). A second ORF, from the opposite end of clone YS11, proved to be identical to 78 bp of a novel gene which was independently identified through a second YAC-based approach. For the latter, DNA from yeast cells containing YAC 14FH1 (FIG. 1B) was digested with EcoRI and subcloned into a plasmid vector. Plasmids that contained human DNA fragments were selected by colony hybridization using total human DNA as a probe. These clones were then used to search for cross-hybridizing sequences as described above 10 for TB1, and the cross-hybridizing clones were subsequently used to screen cDNA libraries. One of the cDNA clones discovered in this way (FH38) contained a long ORF (2496 bp), 78 bp of which were identical to the above-noted sequences in YS11. The ends of the FH38 cDNA clone were 15 then used to initiate cDNA walking to extend the sequence. Eventually, 85 cDNA clones were isolated from normal colon, brain and liver cDNA libraries and found to encompass 8973 nucleotides of contiguous transcript. The gene corresponding to this transcript was named APC. When used 20 as probes for Northern blot analysis, APC cDNA clones hybridized to a single transcript of approximately 9.5 kb, suggesting that the great majority of the gene product was represented in the cDNA clones obtained. Sequences from the 5' end of the APC gene were found in YAC 37HG4 but 25 not in YAC 14FH1. However, the 3' end of the APC gene was found in 14FH1 as well as 37HG4. Analogously, the 5' end of the MCC ceding region was found in YAC clones 19AA9 and 266C3 but not 24ED6 or 14FH1, while the 3 end displayed the opposite pattern. Thus, MCC and APC 30 transcription units pointed in opposite directions, with the direction of transcription going from centromeric to telomeric in the case of MCC, and telomeric to centromeric in the case of APC. PFGE analysis of YAC DNA digested with various restriction endonucleases showed that TB2 and SRP 35 were between MCC and APC, and that the 3' ends of the ceding regions of MCC and APC were separated by approximately 150 kb (FIG. 1B).

Sequence analysis of the APC cDNA clones revealed an open reading frame of 8,535 nucleotides. The 5' end of the 40 ORF contained a methionine codon (codon 1) that was preceded by an in-frame stop codon 9 bp upstream, and the 3' end was followed by several in-frame stop codons. The protein produced by initiation at codon 1 would contain 2,842 amino acids (SEQ ID NO:7) (FIG. 3). The results of 45 database searching with the APC gene product were quite complex due to the presence of large segments with locally biased amino acid compositions. In spite of this, APC could be roughly divided into two domains. The N-terminal 25% of the protein had a high content of leucine residues (12%) 50 and showed local sequence similarities to myosins, various intermediate filament proteins (e.g., desmin, vimentin, neurofilaments) and Drosophila armadillo/human plakoglobin. The latter protein is a component of adhesive junctions (desmosomes) joining epithelial cells (Franke et al., Proc. 55 Natl. Acad. Sci. U.S.A., Vol. 86, p. 4027 (1989); Perfer et al., Cell, Vol. 63, p. 1167 (1990)) The C-terminal 75% of APC (residues 731-2832) is 17% serine by composition with serine residues more or less uniformly distributed. This large domain also contains local concentrations of charged 60 (mostly acidic) and proline residues. There was no indication of potential signal peptides, transmembrane regions, or nuclear targeting signals in APC, suggesting a cytoplasmic localization.

To detect short similarities to APC, a database search was
65 performed using the PAM-40 matrix (Altschul. J. Mol. Bio.,
Vol. 219, p. 555 (1991). Potentially interesting matches to
several proteins were found. The most suggestive of these

involved the ral2 gene product of yeast, which is implicated in the regulation of ras activity (Fukul et al., Mol. Cell. Biol., Vol. 9, p. 5617 (1989)). Little is known about how ral2 might interact with ras but it is interesting to note the positively-charged character of this region in the context of the s negatively-charged GAP interaction region of ras. A specific electrostatic interaction between ras and GAP-related proteins has been proposed.

Because of the proximity of the MCC and APC genes, and the fact that both am implicated in colorectal tumorigenesis, 10 we searched for similarities between the two predicted proteins. Bourne has previously noted that MCC has the potential to form alpha helical coiled coils (Nature, Vol. 351, p. 188 (1991). Lupas and colleagues have recently developed a program for predicting coiled coil potential from 15 primary sequence data (Science, Vol. 252, p. 1162 (1991) and we have used their program to analyze both MCC and APC. Analysis of MCC indicated a discontinuous pattern of coiled-coil domains separated by putative "hinge" or "sparer" regions similar to those seen in laminin and other 20 intermediate filament proteins. Analysis of the APC sequence revealed two regions in the N-terminal domain which had strong coiled coil-forming potential, and these regions corresponded to those that showed local similarities with myosin and IF proteins on database searching. In 25 addition, one other putative coiled coil region was identified in the central region of APC. The potential for both APC and MCC to form coiled coils is interesting in that such structures often mediate homo- and hetero-oligomerization.

Finally, it had previously been noted that MCC shared a short similarity with the region of the m3 muscarinic acetylcholine receptor (mAChR) known to regulate specificity of G-protein coupling. The APC gene also contained a local similarity to the region of the m3 mAChR (SEQ ID NO:9) that overlapped with the MCC similarity (SEQ ID NO:10) (FIG. 4B). Although the similarities to ral2 (SEQ ID NO:8) (FIG. 4A) and m3 mAChR (SEQ ID NO:9) (FIG. 4B) were not statistically significant, they were intriguing in light of previous observations relating G-proteins to neoplasia.

Each of the six genes described above was expressed in normal colon mucosa, as indicated by their representation in colon cDNA libraries. To study expression of the genes in neoplastic colorectal epithelium, we employed reverse transcription-polymerase chain reaction (PCR) assays. Primers based on the sequences of FER, TB1, TB2, MCC, and APC were each used to design primers for PCR performed with cDNA templates. Each of these genes was found to be expressed in normal colon, in each of ten cell lines derived from colorectal cancers, and in tumor cell lines derived from lung and bladder tumors. The ten colorectal cancer cell lines included eight from patients with sporadic CRC and two from patients with FAP.

EXAMPLE 2

This example demonstrates a genetic analysis of the role of the FER gene in FAP and sporadic colorectal cancers.

We considered FER as a candidate because of its proximity to the FAP locus as judged by physical and genetic criteria (see Example 1), and its homology to known 60 tyrosine kinases with oncogenic potential. Primers were designed to PCR-amplify the complete coding sequence of FER from the RNA of two colorectal cancer cell lines derived from FAP patients. cDNA was generated from RNA and used as a template for PCR. The primers used were 65 5'-AGAAGGATCCCTTGTGCAGTGTGGA-3' (SEQ ID NO:95) and 5'-GACAGGATCCTGAAGCTGAGTTTG-3'

(SEQ ID NO:96). The underlined nucleotides were altered from the true FER sequence to create BamHI sites. The cell lines used were JW and Difi, both derived from colorectal cancers of FAP patients. (C. Paraskeva, B. G. Buckle, D.
5 Sheer, C. B. Wigley, Int. J. Cancer 34, 49 (1984); M. E. Gross et al., Cancer Res. 51, 1452 (1991). The resultant 2554 basepair fragments were cloned and sequenced in their entirety. The PCR products were cloned in the BamHI site of Bluescript SK (Stratagene) and pools of at least 50 clones
10 were sequenced en masse using T7 polymerase, as described in Nigro et al., Nature 342, 705 (1989).

Only a single conservative amino acid change (GTG

CTG, creating a val to leu substitution at codon 439) was observed. The region surrounding this codon was then amplified from the DNA of individuals without FAP and this substitution was found to be a common polymorphism, not specifically associated with FAP. Based on these results, we considered it unlikely (though still possible) the FER gene was responsible for FAP. To amplify the regions surrounding codon 439, the following primers were used: 5'-TCAGAAAGTGCTGAAGAG-3' (SEQ ID NO:97) and 5'-GGAATAATTAGGTCTCCAA-3' (SEQ ID NO:98). PCR products were digested with PstI, which yields a 50 bp fragment if codon 439 is leucine, but 26 and 24 bp fragments if it is valine. The primers used for sequencing were chosen from the FER cDNA sequence in Hao et al., supra.

EXAMPLE 3

This example demonstrates the genetic analysis of MCC, TB2, SRP and APC in FAP and sporadic colorectal tumors. Each of these genes is linked and encompassed by contig 3 (see FIG. 1).

Several lines of evidence suggested that this contig was of 35 particular interest. First, at least three of the four genes in this contig were within the deleted region identified in two FAP patients. (See Example 5 infra.) Second, allelic deletions of chromosome 5q21 in sporadic cancers appeared to be centered in this region. (Ashton-Rickardt et al., Oncogene, in press; and Miki et al., Japn. J. Cancer Res., in press.) Some tumors exhibited loss of proximal RFLP markers (up to and potentially including the 5' end of MCC), but no loss of markers distal to MCC. Other tumors exhibited loss of markers distal to and perhaps including the 3' end of 45 MCC, but no loss of sequences proximal to MCC. This suggested either that different ends of MCC were affected by loss in all such cases, or alternatively, that two genes (one proximal to and perhaps including MCC, the other distal to MCC) were separate targets of deletion. Third, clones from 50 each of the six FAP region genes were used as probes on Southern blots containing tumor DNA from patients with Sporadic CRC. Only two examples of somatic changes were observed in over 200 tumors studied: a rearrangement/ deletion whose centromeric end was located within the 55 MCC gene (Kinzler et al., supra) and an 800 bp insertion within the APC gene between nucleotides 4424 and 5584. Fourth, point mutations of MCC were observed in two tumors (Kinzler et al.) supra strongly suggesting that MCC was a target of mutation in at least some sporadic colorectal

Based on these results, we attempted to search for subtle alterations of contig 3 genes in patients with FAP. We chose to examine MCC and APC, rather than TB2 or SRP, because of the somatic mutations in MCC and APC noted above. To facilitate the identification of subtle alterations, the genomic sequences of MCC and APC exons were determined (see Table I, SEQ ID NO:24-38).

TABLE I

APC EXONS

EXON NUCLEOTIDES ¹	EXON BOUNDARY SEQUENCE ²
822 to 930	catgatgttatctgtatttacctatagtctaaattataccatctataatgtgcttaatttttag/GGTTCA(SEQ ID NO: 24)ACCAAG/gtaacagaagattacaaacoctggtcactaatgccatgactacttgctaag (SEO ID NO: 25)
931 to 1309	ggatattaaagtegtaatittigtitetaaacteattiggeecacag/GTGGAA (SEQ ID NO: 26) ATCCAA/gtatgttetetaagtgtaeategtagtgeatg (SEQ ID NO: 27)
1310 to 1405	catcattgctcttcaaataacaaagcattatggtttatgttgattttatttttcag/TGCCAG (SEQ ID NO: 28) AACTAG/gtaagacaaaaagtgtttttatggacatagacaattactggtg (SEQ ID NO: 29)
1406 to 1545	tagatgattgtctttttcctcttgcccttttttaaattag/GGGGAC (SEQ ID NO: 30) AACAAG/gtatgtttttataacatgtatttcttaaggatagctcaggtatga (SEQ ID NO: 31)
1546 to 1623	gettggetteaagttgtetttttaatgateetetattetgtatttaatttaeag/GCTACG (SEQ ID NO: 32) CAGCAG/gtaetatttagaattteaeetgtitttetttittetettittettgaggeagggteteaetetg (SEQ ID. NO: 33)
1624 to 1740	gcaactagtatgattttatgtataaattaatctaaaattgattaatttgacag/GTTATT (SEQ ID NO: 34) AAAAAG/gtacctttgaaaacatttagtactataatatgaattcatgt (SEQ ID NO: 35)
1741 to 1955	caactctaattagatgacccatattcagaaacttactag/GAATCA(SEQ ID NO: 36)CCACAG/gtatatatagagttttatattacttttaaagtacagaattcatactctcaaaaa (SEO ID NO: 37)
1956 to 8973	tettgamtttattteag/GCAAAT (SEQ ID NO: 38) GGTATTTATGCAAAAAAAATGTTTTTGT (SEQ ID NO: 1)

¹Relative to predicted translation initiation site

The entire 3' end of the cloned APC cDNA (nt 1956-8973) appeared to be encoded in this exon, as indicated by restriction endonuclease mapping and sequencing of the cloned genomic DNA. The ORF ended at nt 8535. The extreme 3' end of the APC transcript has not yet been identified.

These sequences were used to design primers for PCR analysis of constitutional DNA from FAP patients.

We first amplified eight exons and surrounding introns of the MCC gene in affected individuals from 90 different FAP kindreds. The PCR products were analyzed by a ribonuclease (RNase) protein assay. In brief, the PCR products were hybridized to in vitro transcribed RNA probes representing the normal genomic sequences. The hybrids were digested with RNase A, which can cleave at single base pair mismatches within DNA-RNA hybrids, and the cleavage 35 products were visualized following denaturing gel electrophoresis. Two separate RNase protection analyses were performed for each exon, one with the sense and one with the antisense strand. Under these conditions, approximately 40% of all mismatches are detectable. Although some amino 40 acid variants of MCC were observed in FAP patients, all such variants were found in a small percentage of normal individuals. These variants were thus unlikely to be responsible for the inheritance of FAP.

We next examined three exons of the A PC gene. The 45 three exons examined included those containing nt 822-930, 931-1309, and the first 300 nt of the most distal exon (nt 1956-2256). PCR and RNase protection analysis were performed as described in Kinzler et al. supra, using the primers underlined in Table I (SEQ ID NO:24-38). The primers for 50 nt 1956-2256 were 5'-GCAAATCCTAAGAGAGAACAA-3' (SEQ ID NO:99) and 5'-GATGGCAAGCITGAGCCAG-3' (SEQ ID NO:100).

In 90 kindreds, the RNase protection method was used to screen for mutations and in an additional 13 kindreds, the 55 PCR products were cloned and sequenced to search for mutations not detectable by RNase protection. PCR products were cloned into a Bluescript vector modified as described in T. A. Holton and M. W. Graham, Nucleic Acids Res. 19, 1156 (1991). A minimum of 100 clones were pooled and 60 sequenced. Five variants were detected among the 103 kindreds analyzed. Cloning and subsequent DNA sequencing of the PCR product of patient P21 indicated a C to T transition in codon 413 that resulted in a change from arginine to cysteine. This amino acid variant was not 65 observed in any of 200 DNA samples from individuals without FAP. Cloning and sequencing of the PCR product

²Small case letters represent introns, large case letters represent exons

from patients P24 and P34, who demonstrated the same abnormal RNase protection pattern indicated that both had a C to T transition at codon 801 that resulted in a change from 30 arginine (CGA) to a stop codon (TGA). This change was not present in 200 individuals without FAP. As this point mutation resulted in the predicted loss of the recognition site for the enzyme Taq I, appropriate PCR products could be digested with Taq I to detect the mutation. This allowed us 35 to determine that the stop codon co-segragated with disease phenotype in members of the family of P24. The inheritance of this change in affected members of the pedigree provides additional evidence for the importance of the mutation.

Cloning and sequencing of the PCR product from FAP patient P93 indicated a C to G transversion at codon 279, also resulting in a stop codon (change from TCA to TGA). This mutation was not present in 200 individuals without FAP. Finally, one additional mutation resulting in a serine 45 (TCA) to stop codon (TGA) at codon 712 was detected in a single patient with FAP (patient P60).

The five germline mutations identified are summarized in Table IIA, as well as four others discussed in Example 9.

5 0			TABLI	E IIA		
	Germline mutations of the APC gene in FAP and GS Patients					
55	EXTRA- COLO- NIC PATIENT DISEASE	CODON	NUCLEO- TIDE CHANGE	AMINO ACID CHANGE	AGE	
60	93 Osteoma	279	TCA->TGA	Ser->Stop	39	Mandi- bular
	24 34	301 301	CGA-> <u>T</u> GA CGA-> <u>T</u> GA	Arg->Stop Arg->Stop	46 27	None Des- moid
65	Tumor 21	413	CGC-> <u>T</u> GC	Arg->Cys	24	Mandi-

bular

TABLE IIA-continued

Gern	nline mutati	ons of the APC gene	in FAP and C	S Patients	_
EXTRA- COLO- NIC PATIENT DISEASE	CODON	NUCLEO- TIDE CHANGE	AMINO ACID CHANGE	AGE	5
Osteoma					10
60 Osteoma	712	TCA->TGA	Ser->Stop	37 Mandi- bular	
3746	243	CAGAG->CAG	splice-		15
3460 3827 3712	301 456 500	CGA-> <u>T</u> GA CTTTCA->CTTCA T-> <u>G</u>	Arg->Stop		

^{*} The mutated nucleotides are underlined.

In addition to these germline mutations, we identified several somatic mutations of MCC and APC in sporadic CRC's. Seventeen MCC exons were examined in 90 sporadic colorectal cancers by RNase protection analysis. In each case where an abnormal RNase protection pattern was observed, 25 the corresponding PCR products were cloned and sequenced. This led to the identification of six point mutations (two described previously) (Kinzler et al., supra), each of which was not found in the germline of these patients (Table IIB).

TABLE IIB

	Somatic Mutations in Sporadic CRC Patients				
PATIENT	CODON ¹	NUCLEOTIDE CHANGE	AMINO ACID CHANGE	35	
T35	MCC 12	GAG/gtaaga-> GAG/gtaaaa	(Splice Donor)	_	
T16	MCC 145	ctcag/GGA-> atcag/GGA	(Splice Acceptor)	40	
T47	MCC 267	ČGG->CTG	Arg->Leu		
T81	MCC 490	TCG->TTG	Ser->Leu		
T35	MCC 506	CGG->CAG	Arg->Gln		
T 91	MCC 698	GCT->GT	Ala->Val		
T34	APC 288	CCAGT->CCCAGCCAGT	(Insertion)		
T27	APC 331	CGA->TGA	Arg->Stop	45	
T135	APC 437	CAA/gtaa->CAA/gcaa	(Splice Donor)		
T20I	APC 1338	CAG-> <u>T</u> AG	Gln->Stop		

For splice site mutations, the codon nearest to the mutation is listed. The underlined nucleotides were mutant; small case letters represent introns, 50 large case letters represent exons

Four of the mutations resulted in amino acid substitutions and two resulted in the alteration of splice site consensus elements. Mutations at analogous splice site positions in other genes have been shown to alter RNA processing in 55 vivo and in vitro.

Three exons of APC were also evaluated in sporadic tumors. Sixty tumors were screened by RNase protection, and an additional 98 tumors were evaluated by sequencing. The exons examined included nt 822–930, 931–1309, and 60 1406–1545 (Table I). A total of three mutations were identified, each of which proved to be somatic. Tumor T27 contained a somatic mutation of CGA (arginine) to TGA (stop codon) at codon 33. Tumor T135 contained a GT to GC change at a splice donor site. Tumor T34 contained a 5 bp 65 insertion (CAGCC between codons 288 and 289) resulting in a stop at codon 291 due to a frameshift.

We serendipitously discovered one additional somatic mutation in a colorectal cancer. During our attempt to define the sequences and splice patterns of the MCC and APC gene products in colorectal epithelial cells, we cloned cDNA from 5 the colorectal cancer cell line SW480. The amino acid sequence of the MCC gene from SW480 was identical to that previously found in clones from human brain. The sequence of APC in SW480 cells, however, differed significantly, in that a transition at codon 1338 resulted in a 10 change from glutamine (CAG) to a stop codon (TAG). To determine if this mutation was somatic, we recovered DNA from archival paraffin blocks of the original surgical specimen (T201) from which the tumor cell line was derived 28 years ago.

DNA was purified from paraffin sections as described in S. E. Goelz, S. R. Hamilton, and B. Vogelstein. Biochem. Biophys. Res. Comm. 130, 118 (1985). PCR was performed as described in reference 24, using the primers 5'-GITCCAGCAGTGTCACAG-3' (SEQ ID NO:101) and
 5'-GGGAGATTTCGCTCCTGA-3' (SEQ ID NO:102). A PCR product containing codon 1338 was amplified from the archival DNA and used to show that the stop codon represented a somatic mutation present in the original primary tumor and in cell lines derived from the primary and metastatic tumor sites, but not from normal tissue of the patient.

The ten point mutations in the MCC and APC genes so far discovered in sporadic CRCs are summarized in Table IIB. Analysis of the number of mutant and wild-type PCR clones obtained from each of these tumors showed that in eight of the ten cases, the wild-type sequence was present in approximately equal proportions to the mutant. This was confirmed by RFLP analysis using flanking markers from chromosome 5q which demonstrated that only two of the ten tumors (T135 and T201) exhibited an allelic deletion on chromosome 5q. These results are consistent with previous observations showing that 20-40% of sporadic colorectal tumors had allelic deletions of chromosome 5q. Moreover, these data suggest that mutations of 5q21 genes are not limited to those colorectal tumors which contain allelic deletions of this chromosome.

EXAMPLE 4

This example characterizes small, nested deletions in DNA from two unrelated FAP patients.

DNA from 40 FAP patients was screened with cosmids that has been mapped into a region near the APC locus to identify small deletions or rearrangements. Two of these cosmids, L5.71 =nd L5.79, hybridized with a 1200 kb NotI fragment in DNAs from most of the FAP patients screened.

The DNA of one FAP patient, 3214, showed only a 940 kb NotI fragment instead of the expected 1200 kb fragment. DNA was analyzed from four other members of the patient's immediate family; the 940 kb fragment was present in her affected mother (4711), but not in the other, unaffected family members. The mother also carried a normal 1200 kb NotI fragment that was transmitted to her two unaffected offspring. These observations indicated that the mutant polyposis allele is on the same chromosome as the 940 kb NotI fragment. A simple interpretation is that APC patients 3214 and 4711 each carry a 260 kb deletion within the APC locus.

If a deletion were present, then other enzymes might also 65 be expected to produce fragments with altered mobilities. Hybridization of L5.79 to NruI-digested DNAs from both affected members of the family revealed a novel NruI fragment of 1300 kb, in addition to the normal 1200 kb NruI fragment. Furthermore, MluI fragments in patients 3214 and 4711 also showed an increase in size consistent with the deletion of an MluI site. The two chromosome 5 homologs of patient 3214 were segregated in somatic cell hybrid lines; 5 HHW1155 (deletion hybrid) carried the abnormal homolog and HHW1159 (normal hybrid) carried the normal homolog.

Because patient 8214 showed bray a 940 kb NotI fragment, she had not inherited the 1200 kb fragment present in the unaffected father's DNA. This observation suggests 10 that he must be heterozygous for, and have transmitted, either a deletion of the L5.79 probe region or a variant NotI fragment too large to resolve on the gel system. As expected, the hybrid cell line HHW1159, which carries the paternal homolog, revealed no resolved Not fragment when probed with L5.79. However, probing of HHW1159 DNA with 15 L5.79 following digestion with other enzymes did reveal restriction fragments, demonstrating the presence of DNA homologous to the probe. The father is, therefore, interpreted as heterozygous for a polymorphism at the NotI site, with one chromosome 5 having a 1200 kb NotI fragment and 20 the other having a fragment too large to resolve consistently on the gel. The latter was transmitted to patient 3214.

When double digests were used to order restriction sites within the 1200 kb NotI fragment, L5.71 and L5.79 were beth found to lie on a 550 kb NotI-NruI fragment and, 25 therefore, on the same side of an NruI site in the 1200 kb NotI fragment. To obtain genomic representation of sequences present over the entire 1200 kb Notl fragment, we constructed a library of small-fragment inserts enriched for sequences from this fragment. DNA from the somatic cell 30 hybrid HHW141, which contains about 40% of chromosome 5, was digested with NotI and electrophoresed under pulsedfield gel (PFG) conditions; EcoRI fragments from the 1200 kb region of this gel were cloned into a phage vector. Probe Map30 was isolated from this library. In normal individuals 35 probe Map30 hybridizes to the 1200 kb NotI fragment and to a 200 kb NruI fragment. This latter hybridization places Map30 distal, with respect to the locations of L5.71 and L5.79, to the NruI site of the 550 kb NotI-NruI fragment.

Because Map30 hybridized to the abnormal, 1300 kb Nrul fragment of patient 3214, the locus defined by Map30 lies outside the hypothesized deletion. Furthermore, in normal chromosomes Map30 identified a 200 kb NruI fragment and L5.79 identified a 1200 kb NruI fragment; the hypothesized deletion must, therefore, be removing an NruI site, or sites, lying between Map30 and L5.79, and these two probes must flank the hypothesized deletion. A restriction map of the genomic region, showing placement of these probes, is shown in FIG. 5.

A NotI digest of DNA from another FAP patient, 3824, 50 was probed with L5.79. In addition to the 1200 kb normal NotI fragment, a fragment of approximately 1100 kb was observed, consistent with the presence of a 100 kb deletion in one chromosome 5. In this case, however, digestion with NruI and MluI did not reveal abnormal bands, indicating that 55 if a deletion were present, its boundaries must lie distal to the NruI and MluI sites of the fragments identified by L5.79. Consistent with this expectation, hybridization of Map30 to DNA from patient 3824 identified a 760 kb MluI fragment in addition to the expected 860 kb fragment, supporting the interpretation of a 100 kb deletion in this patient. The two chromosome 5 homologs of patient 3824 were segregated in somatic cell hybrid lines; HHW1291 was found to carry only the abnormal homolog and HHW1290 only the normal homolog.

That the 860 kb Mlul fragment identified by Map30 is distinct from the 830 kb Mlul fragment identified previously

by L5.79 was demonstrated by hybridization of Map30 and L5.79 to a NotI-MluI double digest of DNA from the hybrid cell (HHW1159) containing the nondeleted chromosome 5 homolog of patient 3214. As previously indicated, this 5 hybrid is interpreted as missing one of the NotI sites that define the 1200 kb fragment. A 620 kb NotI-MluI fragment was seen with probe L5.79, and an 860 kb fragment was seen with Map30. Therefore, the 830 kb MluI fragment recognized by probe L5.79 must contain a NotI site in HHW1159 DNA; because the 860 kb MluI fragment remains intact, it does not carry this NotI site and must be distinct from the 830 kb MluI fragment.

EXAMPLE 5

This example demonstrates the isolation of human sequences which span the region deleted in the two unrelated FAP patients characterized in Example 4.

A strong prediction of the hypothesis that patients 8214 and 3824 carry deletions is that some sequences present on normal chromosome 5 homologs would be missing from the hypothesized deletion homologs. Therefore, to develop genomic probes that might confirm the deletions, as well as to identify genes from the region, YAC clones from a contig seeded by cosmid L5.79 were localized from a library containing seven haploid human genome equivalents (Albertsen et al., Proc. Natl. Acad. Sci. U.S.A., Vol. 87, pp. 4256-4260 (1990)) with respect to the hypothesized deletions. Three clones, YACs 57B8, 310D8, and 183H12, were found to overlap the deleted region.

Importantly, one end of YAC 57B8 (clone AT57) was found to lie within the patient 3214 deletion. Inverse polymerase chain reaction (PCR) defined the end sequences of the insert of YAC 57B8. PCR primers based on one of these end sequences repeatedly failed to amplify DNA from the somatic cell hybrid (HHW1155) carrying the deleted homolog of patient 3214, but did amplify a product of the expected size from the somatic cell hybrid (HHW1159) carrying the normal chromosome 5 homolog. This result supported the interpretation that the abnormal restriction fragments found in the DNA of patient 3214 result from a deletion.

Additional support for the hypothesis of deletion in DNA from patient 3214 came from subcloned fragments of YAC 183H12, which spans the region in question. Y11, an EcoRI fragment cloned from YAC 183H12, hybridized to the normal, 1200 kb NotI fragment of patient 4711, but failed to hybridize to the abnormal, 940 kb NotI fragment of 4711 or to DNA from deletion cell line HHW1155. This result confirmed the deletion in patient 3214.

Two additional EcoRl fragments from YAC 183H12, Y10 and Y14, were localized within the patient 3214 deletion by their failure to hybridize to DNA from HHW1155. Probe Y10 hybridizes to a 150 kb NruI fragment in normal 55 chromosome 5 homologs. Because the 3214 deletion creates the 1300 kb NruI fragment seen with the probes L5.79 and Map30 that flank the deletion, these NruI sites and the 150 kb NruI fragment lying between must be deleted in patient 3214. Furthermore, probe Y10 hybridizes to the same 620 kb Notl-MluI fragment seen with probe L5.79 in normal DNA, indicating its location as L5.79-proximal to the deleted MluI site and placing it between the MluI site and the L5.79-proximal NruI site. The MluI site must, therefore, lie between the NruI sites that define the 150 kb NruI fragment 65 (see FIG. 5).

Probe Y11 also hybridized to the 150 kb NruI fragment in the normal chromosome 5 homolog, but failed to hybridize to the 620 kb Notl-MluI fragment, placing it L5.79-distal to the MluI site, but proximal to the second NruI site. Hybridization to the same (860 kb) MluI fragment as Map30 confirmed the localization of probe Y11 L5.79-distal to the MluI site.

Probe Y14 was shown to be L5.79-distal to both deleted NruI sites by virtue of its hybridization to the same 200 kb NruI fragment of the normal chromosome 5 seen with Map30. Therefore, the order of these EcoRI fragments derived from YAC 183H12 and deleted in patient 3214, with respect to L5.79 and Map30, is L5.79-Y10-Y11-Y14-Map30.

The 100 kb deletion of patient 3824 was confirmed by the failure of aberrant restriction fragments in this DNA to hybridize with probe Y11, combined with positive hybrid-15 izations to probes Y10 and/or Y14. Y10 and Y14 each hybridized to the 1100 kb NotI fragment of patient 3824 as well as to the normal 1200 kb NotI fragment, but Y11 hybridized to the 1200 kb fragment only. In the Mlul digest, probe Y14 hybridized to the 860 kb and 760 kb fragments of patient 3824 DNA, but probe Y11 hybridized only to the 860 k13 fragment. We conclude that the basis for the alteration in fragment size in DNA from patient 3824 is, indeed, a deletion. Furthermore, because probes Y10 and Y14 are missing from the deleted 3214 chromosome, but present on the deleted 3824 chromosome, and they have been shown to flank probe Y11, the deletion in patient 3824 must be nested within the patient 3214 deletion.

Probes Y10, Y11, Y14 and Map30 each hybridized to YAC 310D8, indicating that this YAC spanned the patient 3824 deletion and at a minimum, most of the 3214 deletion. The YAC characterizations, therefore, confirmed the presence of deletions in the patients and provided physical representation of the deleted region.

EXAMPLE 6

This example demonstrates that the MCC coding sequence maps outside of the region deleted in the two FAP patients characterized in Example 4.

An intriguing FAP candidate gene, MCC, recently was ascertained with cosmid L5.71 and was shown to have undergone mutation in colon carcinomas (Kinzler et al., supra). It was therefore of interest to map this gene with respect to the deletions in APC patients. Hybridization of 45 MCC probes with an overlapping series of YAC clones extending in either direction from L5.71 showed that the 3' end of MCC must be oriented toward the region of the two APC deletions.

Therefore, two 3' cDNA clones from MCC were mapped 50 with respect, to the deletions: clone 1CI (bp 2378-4181) and clone 7 (bp 2890-3560). Clone 1CI contains sequences from the C-terminal end of the open reading frame, which stops at nucleotide 2708, as well as 3' untranslated sequence. Clone 7 contains sequence that is entirely 3' to the open 55 reading frame. Importantly, the entire 3' untranslated sequence contained in the cDNA clones consists of a single 2.5 kb exon. These two clones were hybridized to DNAs from the YACs spanning the FAP region. Clone 7 fails to hybridize to YAC 310D8, although it does hybridize to 60 YACs 183H12 and 57B8; the same result was obtained with the cDNA 1CL Furthermore, these probes did show hybridization to DNAs from both hybrid cell lines (HWW1159 and HWW1155) and the lymphoblastoid cell line from patient 3214, confirming their locations outside the deleted region. 65 Additional mapping experiments suggested that the 3' end of the MCC cDNA clone contig is likely to be located more

than 45 kb from the deletion of patient 3214 and, therefore, more than 100 kb from the deletion of patient 3824.

EXAMPLE 7

This example identifies three genes within the deleted region of chromosome 5 in the two unrelated FAP patients characterized in Example 4.

Genomic clones were used to screen cDNA libraries in 10 three separate experiments. One screening was done with a phage clone derived from YAC 310D8 known to span the 260 kb deletion of patient 3214. A large-insert phage library was constructed from this YAC; screening with Y11 identified \$205, which mapped within both deletions. When 15 clone λ205 was used to probe a random-, plus oligo(dT)-, primed fetal brain cDNA library (approximately 300,000 phage), six cDNA clones were isolated and each of them mapped entirely within both deletions. Sequence analysis of these six clones formed a single cDNA contig, but did not 20 reveal an extended open reading frame. One of the six cDNAs was used to isolate more cDNA clones, some of which crossed the L5.71-proximal breakpoint of the 3824 deletion, as indicated by hybridization to both chromosome of this patient. These clones also contained an open reading 25 frame, indicating a transcriptional orientation proximal to distal with respect to L5.71. This gene was named DP1 (deleted in polyposis 1). This gene is identical to TB2 described above.

cDNA walks yielded a cDNA contig of 3.0-3.5 kb, and included two clones containing terminal poly(A) sequences. This size corresponds to the 3.5 kb band seen by Northern analysis. Sequencing of the first 3163 bp of the cDNA contig revealed an open reading frame extending from the first base to nucleotide 631, followed by a 2.5 kb 3' untranslated region. The sequence surrounding the methionine codon at base 77 conforms to the Kozak consensus of an initiation methionine (Kozak, 1984). Failed attempts to walk farther, coupled with the similarity of the lengths of isolated cDNA and mRNA, suggested that the NH₂-terminus of the DP1 protein had been reached. Hybridization to a combination of genomic and YAC DNAs cut with various enzymes indicated the genomic coverage of DP1 to be approximately 30

Two additional probes for the locus, YS-11 and YS-39, 45 which had been ascertained by screening of a cDNA library with an independent YAC probe identified with MCC sequences adjacent to L5.71, were mapped into the deletion region. YS-39 was shown to be a cDNA identical in sequence to DP1. Partial characterization of YS-11 had 50 shown that 200 bp of DNA sequence at one end was identical to sequence coding for the 19 kd protein of the ribosomal signal recognition particle, SRP19 (Lingelbach et al., supra). Hybridization experiments mapped YS-11 within beth deletions. The sequence of this clone, however, was 55 found to be complex. Although 454 bp of the 1032 bp sequence of YS-11 were identical to the GenBank entry for the SRP19 gene, another 578 bp appended 5' to the SRP19 sequence was found to consist of previously unreported sequence containing no extended open reading frames. This 60 suggested that YS-11 was either a chimeric clone containing two independent inserts or a clone of an incompletely processed or aberrant message. If YS-11 were a conventional chimeric clone, the independent segments would not be expected to map to the same physical region. The 65 segments resulting from anomalous processing of a continuous transcript, however, would map to a single chromosomal region.

Inverse PCR with primers specific to the two ends of YS-11, the SRP19, end and the unidentified region, verified that both sequences map within the YAC 310D8; therefore, YS-11 is most likely a clone of an immature or anomalous mRNA species. Subsequently, both ends were shown to lie 5 with the deleted region of patient 3824, and YS-11 was used to screen for additional cDNA clones.

Of the 14 cDNA clones selected from the fetal brain library, one clone, V5, was of particular interest in that it contained an open reading frame throughout, although it ¹⁰ included only a short identity to the first 78 5' bases of the YS-11 sequence. Following the 78 bp of identical sequence, the two cDNA sequences diverged at an AG. Furthermore, divergence from genomic sequence was also seen after these 78 bp, suggesting the presence of a splice junction, and ¹⁵ supporting the view that YS-11 represents an irregular message.

Starting with V5, successive 5' and 3' walks were performed; the resulting cDNA contig consisted of more than 100 clones, which defined a new transcript, DP2. Clones walking in the 5' direction crossed the 3824 deletion breakpoint farthest from L5.71; since its 3' end is closer to this cosmid than its 5' end, the transcriptional orientation of DP2 is opposite to that of MCC and DP1.

The third screening approach relied on hybridization with a 120 kb MluI fragment from YAC 57B8. This fragment hybridizes with probe Y11 and completely spans the 100 kb deletion in patient 3824. the fragment was purified on two preparative PFGs, labeled, and used to screen a fetal brain cDNA library. A number of cDNA clones previously identified in the development of the DP1 and DP2 contigs were reascertained. However, 19 new cDNA clones mapped into the patient 3824 deletion. Analysis indicated that these 19 formed a new contig, DP3, containing a large open reading 35 frame.

A clone from the 5' end of this new cDNA contig hybridized to the same EcoRI fragment as the 3' end of DP2. Subsequently, the DP2 and DP3 contigs were connected by a single 5' walking step from DP3, to form the single contig 40 DP2.5. The complete nucleotide sequence of DP2.5 is shown in FIG. 9.

The consensus cDNA sequence of DP2.5 suggests that the entire coding sequence of DP2.5 has been obtained and is 8532 bp long. The most 5' ATG codon occurs two codons from an in-frame stop and conforms to the Kozak initiation 5 consensus (Kozak, Nucl. Acids. Res., Vol. 12, p. 857–872 1984). The 3' open reading frame breaks down over the final 1.8 kb, giving multiple stops in all frames. A poly(A) sequence was found in one clone approximately 1 kb into the 3' untranslated region, associated with a polyadenylation signal 33 bp upstream (position 9530). The open reading frame is almost identical to that identified as APC above.

An alternatively spliced exon at nucleotide 934 of the DP2.5 transcript is of potential interest. it was first discovered by noting that two classes of cDNA had been isolated.

The more abundant cDNA class contains a 303 bp exon not included in the other. The presence in vivo of the two transcripts was verified by an exon connection experiment. Primers flanking the alternatively spliced exon were used to amplify, by PCR, cDNA prepared from various adult tissues.

Two PCR products that differed in size by approximately 300 bases were amplified from all the tissues tested; the larger product was always more abundant than the smaller.

EXAMPLE 8

This example demonstrates the primers used to identify subtle mutations in DP1, SRP19, and DP25.

To obtain DNA sequence adjacent to the exons of the genes DP1, DP2.5, and SRP19, sequencing substrate was obtained by inverse PCR amplification of DNAs from two YACs, 310D8 and 183H12, that span the deletions. Ligation at low concentration cyclized the restriction enzymedigested YAC DNAs. Oligonucleotides with sequencing tails, designed in inverse orientation at intervals along the cDNAs, primed PCR amplification from the cyclized templates. Comparison of these DNA sequences with the cDNA sequences placed exon boundaries at the divergence points. SRP19 and DP1 were each shown to have five exons. DP2.5 consisted of 15 exons. The sequences of the oligonucleotides synthesized to provide PCR amplification primers for the exons of each of these genes are listed in Table III SEQ ID NO:39-94.

TABLE III

Sequences of Primers Used for SSCP Analyses						
Exon	Primer 1	Primer 2				
	1	DP1				
	UP-TOCCOGCCTGCCGCTCTC UP-GTGAACGGCTCTCATGCTGC UP-ATGATATCTTACCAAATGATATAC	RP-GCAGCGGCGGCTCCCGTG RP-ACCTGCGGGGGGGAATGGA RP-TTATTCCTACTTCTTCTATACAG				
	UP-TACCCATGCTGGCTCTTTTTC	RP-TGGGGCCATCTTGTTCCTGA				

UP-TGCGGCTCCTGGGTTGTTG
UP-TTTTCTCCTGCCTCTTACTGC
UP-CCACTTAAAGCACATATATTTAGT
UP-TTCTTAAGTCCTGTTTTTCTTTTG
UP-CTCAGATTATACACTAAGCCTAAC

UP-ACATTAGGCACAAAGCTTGCAA

RP-GCCCCTTCCTTTCTGAGGAC RP-ATGACACCCCCCATTCCCTC RP-GTATGGAAAATAGTGAAGAACC RP-TTTAGAACCTTTTTTGTGTTGTG RP-CATGTCTCTTACAGTAGTACCA

RP-ATCAAGCTCCAGTAAGAAGGTA

DP2.5

SRP19

UP-AGGTCCAAGGGTAGCCAAGG*
UP-AAATACAGAATCAIGICTTGAAGT
UP-TAACTTAGATAGCAGTAATTTCCC*
UP-ATAGGTCATTGCTTTTACTGATTAACG
UP-GTAGCCATAGTATGATTAACG
UP-GGTAGCCATAGTATGATTATTTCT

RP-TAAAAATGGATAAACTACAATTAAAAG RP-ACACCTAAAGATGACAATTIGAG RP-ACAATAAACTGGAGTACACAAGG RP-TGAATTTTAATGGATTACCTAGGT RP-TGTAATTCATTTTATTCCTAATACCTC RP-CTACCTATTTTTATACCCACAAAC

TABLE III-continued

Sequences of Primers Used for SSCP Analyses

Exon	Primer 1	Primer 2
	UP-AAGAAAGCCTACACCATTTTTGC	RP-GATCATTCTTAGAACCATCTTGC
	UP-ACCTATAGTCTAAATTATACCATC	RP-GTCATGGCATTACTGACCAG
	UP-AGTCGTAATTTTGTTTCTAAACTC	RP-TGAAGGACTCCGATTTCACCC*
	UP-TCATTCACTCACAGCCTGATGAC*	RP-GCTTTGAAACATGCACTACGAT
	UP-AAACATCATTGCTCTTCAAATAAC	RP-TACCATGATTTAAAAATCCACCAG
	UP-GATGATTGTCTTTTTCCTCTTTGC	RP-CTGAGCTATCTTAAGAAATACATG
	UP-TITTAAATGATCCTCTATTCTGTAT	RP-ACAGAGTCAGACCCTCCCTCAAAG
	UP-TTTCTATTCTTACTGCTAGCATT	RP-ATACACAGGTAAGAAATTAGGA
	UP-TAGATGACCCATATTCTCTTTC	RP-CAATTAGGTCTTTTTGAGAGTA
3-A	UP-GTTACTGCATACACATTGTGAC	RP-GCTTTTTGTTTCGTAACATGAAG*
-B	UP-AGTACAAGGATGCCAATATTATG*	RP-ACTICIATCTTTTTCAGAACGAG*
-c	UP-ATTTGAATACTACAGTGTTACCC*	RP-CTTGTATTCTAATTTGGCATAAGG*
-D	UP-CTGCCCATACACATTCAAACAC*	RP-TGTTTGCGTCTTGCCCATCTT*
-E	UP-AGTCTTAAATATTCAGATGAGCAG*	RP-GITTCTCTTCATTATATTTTATGCTA*
-F	UP-AAGCCTACCAATTATAGTGAACG*	RP-AGCTGATGACAAAGATGATAATC*
-G	UP-AAGAAACAATACAGACTTATTGTG*	RP-ATGAGTGGGGTCTCCTGAAC*
-H	UPATCTCCCTCCAAAAGTGGTGC*	RP-TCCATCTGGAGTACTTTCTGTG*
-I	UP-AGTAAATGCTGCAGTTCAGAGG*	RP-CCGTGGCATATCATCCCCC*
-J	UP-CCCAGACTGCTTCAAAATTACC*	RP-GAGCCTCATCTGTACTTCTGC*
-K	UP-CCCTCCAAATGAGTTAGCTGC*	RP-TTGTGGTATAGGTTTTACTGGTG*
-L	UP-ACCCAACAAAAATCAGTTAGATG*	RP-GTGGCTGGTAACTTTAGCCTC*
-N	UP-ATGATGTTGACCTTTCCAGGG*	RP-ATTGTGTAACTTTTCATCAGTTGC*
-M	UP-AAAGACATACCAGACAGAGGG*	RP-CTTTTTTGGCATTGCGGAGCT*
-0	UP-AAGATGACCTGTTGCAGGAATG*	RP-GAATCAGACCAAGCTTGTCTAGAT*
-P	UP-CAATAGTAAGTAGTTTACATCAAG*	RP-AAACAGGACTTGTACTGTAGGA*
-Q	UP-CAGCCCCTTCAAGCAAACATC*	RP-GAGGACTTATTCCATTTCTACC*
-R	UP-CAGTCTCCTGGCCGAAACTC*	RP-GTTGACTGGCGTACTAATACAG*
-S	UP-TGGTAATGGAGCCAATAAAAAGG*	RP-TGGGACTTTTCGCCATCCAC*
-T	UP-TGTCTCTATCCACACATTCGTC*	RP-ATGITTTTCATCCTCACTITTTGC*
-U	UP-GGAGAAGAACTGGAAGTTCATC*	RP-TTGAATCTTTAATGTTTGGATTTGC*
-V	UP-TCTCCCACAGGTAATACTCCC	RP-GCTACAACTGAATGGGGTACG
-₩	UP-CAGGACAAAATAATCCTGTCCC	RP-ATTITCTTACTTTCATTCTTCCTC

All primers are read in the 5' to 3' direction, the first primer in each pair lies 5' of the exon it amplifies: the second primer lies 3' of the exon it amplifies. Primers that lie within the exon are identified by an asterisk. UP represents the -21M13 universal primer sequence:

RP represents the M13 reverse primer sequence.

With the exception of exons 1, 3, 4, 9, and 15 of DP2.5 (see below), the primer sequences were located in intron sequences flanking the exons. The 5' primer of exon 1 is 40 complementary to the cDNA sequence, but extends just into the 5' Kozak consensus sequence for the initiator methionine, allowing a survey of the translated sequences. The 5' primer of exon 3 is actually in the 5' coding sequences of this exon, as three separate intronic primers simply would 45 not amplify. The 5' primer of exon 4 just overlaps the 5' end of this exon, and we thus fail to survey the 19 most 5' bases of this exon. For exon 9, two overlapping primer sets were used, such that each had one end within the exon. For exon 15, the large 3' exon of DP2.5, overlapping primer pairs 50 were placed along the length of the exon; each pair amplified a product of 250-400 bases.

EXAMPLE 9

This example demonstrates the use of single stranded 55 conformation polymorphism (SSCP) analysis as described by Orita et al. Proc. Natl. Acad. Sci. U.S.A., Vol. 86, pp. 2766-70 (1989) and Genomics, Vol. 5, pp. 874-879 (1989) as applied to DP1, SRP19 and DP2.5.

SSCP analysis identifies most single- or multiple-base 60 changes in DNA fragments up to 400 bases in length. Sequence alterations are detected as shifts in electrophoretic mobility of single-stranded DNA on nondenaturing acrylamide gels; the two complementary strands of a DNA segment usually resolve as two SSCP conformers of distinct 65 mobilities. However, if the sample is from an individual heterozygous for a base-pair variant within the amplified

segment, often three or more bands are seen. In some cases, even the sample from a homozygous individual will show multiple bands. Base-pair-change variants are identified by differences in pattern among the DNAs of the sample set.

Exons of the candidate genes were amplified by PCR from the DNAs of 61 unrelated FAP patients and a control set of 12 normal individuals. The five exons from DP1 revealed no unique conformers in the FAP patients, although common conformers were observed with exons 2 and 3 in some individuals of both affected and control sets, indicating the presence of DNA sequence polymorphisms. Likewise, none of the five exons of SRP19 revealed unique conformers in DNA from FAP patients in the test panel.

Testing of exons 1 through 14 and primer sets A through N of exon 15, of the DP2.5 gene, however, revealed variant conformers specific to FAP patients in exons 7, 8, 10, 11, and 15. These variants were in the unrelated patients 3746, 3460, 3827, 3712, and 3751, respectively. The PCR-SSCP proce-55 dure was repeated for each of these exons in the five affected individuals and in an expanded set of 48 normal controls. The variant bands were reproducible in the FAP patients but were not observed in any of the control DNA samples. Additional variant conformers in exons 11 and 15 of the DP2.5 gene were seen; however, each of these was found in both the affected and control DNA sets. The five sets of conformers unique to the FAP patients were sequenced to determine the nucleotide changes responsible for their altered mobilities. The normal conformers from the host individuals were sequenced also. Bands were cut from the

65 dried acrylamide gels, and the DNA was eluted. PCR amplification of these DNAs provided template for sequencing.

The sequences of the unique conformers from exons 7, 8, 10, and 11 of DP2.5 revealed dramatic mutations in the DP2.5 gene. The sequence of the new mutation creating the exon 7 conformer in patient 3746 was shown to contain a deletion of two adjacent nucleotides, at positions 730 and 5 731 in the cDNA sequence (FIG. 7, SEQ ID NO:1). The normal sequence at this splice junction is CAGGGTCA (intronic sequence underlined), with the intron-exon boundary between the two repetitions of AG. The mutant allele in this patient has the sequence CAGGTCA. Although this change is at the 5' splice site, comparison with known consensus sequences of splice junctions would suggest that a functional splice junction is maintained. If this new splice junction were functional, the mutation would introduce a frameshift that creates a stop codon 15 nucleotides downstream. If the new splice junction were not functional. messenger processing would be significantly altered.

To confirm the 2-base deletion, the PCR product from FAP patient 3746 and a control DNA were electrophoresed on an acrylamide-urea denaturing gel, along with the products of a sequencing reaction. The sample from patient 3746 showed two bands differing in size by 2 nucleotides, with the larger band identical in mobility to the control sample; this result was independent confirmation that patient, 3746 is heterozygous for a 2 bp deletion.

The unique conformer found in exon 8 of patient 3460 was found to carry a C-T transition, at position 904 in the cDNA sequence of DP2.5 (shown in FIG. 7), which replaced the normal sequence of CGA with TGA. This point mutation, when read in frame, results in a stop codon replacing the normal arginine codon. This single-base change had occurred within the context of a CG dimer, a potential hot spot for mutation (Barker et al., 1984).

The conformer unique to FAP patient 3827 in exon 10 was found to contain a deletion of one nucleotide (1367, 1368, or 1369) when compared to the normal sequence found in the other bands on the SSCP gel. This deletion, occurring within a set of three T's, changed the sequence from CTTTCA to CTTCA; this 1 base frameshift creates a downstream stop within 30 bases. The PCR product amplified from this patient's DNA also was electrophoresed on an acrylamide-urea denaturing gel, along with the PCR product from a control DNA and products from a sequencing reaction. The patient's PCR product showed two bands differing by 1 bp in length, with the larger identical in mobility to the PCR product from the normal DNA; this result confirmed the presence of a 1 bp deletion in patient 3827.

Sequence analysis of the variant conformer of exon 11 from patient 3712 revealed the substitution of a T by a G at position changing the normal tyrosine codon to a stop codon. 50

The pair of conformers observed in exon 15 of the DP2.5 gene for FAP patient 3751 also was sequenced. These conformers were found to carry a nucleotide substitution of C to G at position 5253, the third base of a valine codon. No amino acid change resulted from this substitution, suggesting that this conformer reflects a genetically silent polymorphism.

The observation of distinct inactivating mutations in the DP2.5 gene in four unrelated patients strongly suggested that DP2.5 is the gene involved in FAP. These mutations are summarized in Table IIA.

EXAMPLE 10

This example demonstrates that the mutations identified in the DP2.5 (APC) gene segregate with the FAP phenotype. 65 Patient 3746, described above as carrying an APC allele with a frameshift mutation, is an affected offspring of two

normal parents. Colonoscopy revealed no polyps in either parent nor among the patient's three siblings.

DNA samples from both parents, from the patient's wife, and from their three children were examined. SSCP analysis of DNA from both of the patient's parents displayed the normal pattern of conformers for exon 7, as did DNA from the patients's wife and one of his off-spring. The two other children, however, displayed the same new conformers as their affected father. Testing of the patient and his parents with highly polymorphic VNTR (variable number of tandem repeat) markers showed a 99.98% likelihood that they are his biological parents.

These observations confirmed that this novel conformer, known to reflect a 2 bp deletion mutation in the DP2.5 gene, appeared spontaneously with FAP in this pedigree and was transmitted to two of the children of the affected individual.

EXAMPLE 11

This example demonstrates polymorphisms in the APC gene which appear to be unrelated to disease (FAP).

Sequencing of variant conformers found among controls as well as individuals with APC has revealed the following polymorphisms in the APC gene: first, in exon 11, at position 1458, a substitution of T to C creating an RsaI restriction site but no amino acid change; and second, in exon 15, at positions 5037 and 5271, substitutions of A to G and G to T, respectively, neither resulting in amino acid substitutions. These nucleotide polymorphisms in the APC gene sequence may be useful for diagnostic purposes.

EXAMPLE 12

This example shows the structure of the APC gene.

The structure of the APC gene is schematically shown in FIG. 8, with flanking intron sequences indicated (SEQ ID NO:11-38).

The continuity of the very large (6.5 kb), most 3' exon in DP2.5 was shown in two ways. First, inverse PCR with primers spanning the entire length of this exon revealed no divergence of the cDNA sequence from the genomic sequence. Second, PCR amplification with converging primers placed at intervals along the exon generated products of the same size whether amplified from the originally isolated cDNA, cDNA from various tissues, or genomic template. Two forms of exon 9 were found in DP2.5: one is the complete exon; and the other, labeled exon 9A, is the result of a splice into the interior of the exon that deletes bases 934 to 1236 in the mRNA and removes 101 amino acids from the predicted protein (see FIG. 3, SEQ ID NO:1 and 2).

EXAMPLE 13

This example demonstrates the mapping of the FAP deletions with respect to the APC exons.

Somatic cell hybrids carrying the segregated chromosomes 5 from the 100 kb (HHW1291) and 260 kb (HHW1155) deletion patients were used to determine the distribution of the APC genes exons across the deletions. DNAs from these cell lines were used as template, along with genomic DNA from a normal control, for PCR-based amplification of the APC exons.

PCR analysis of the hybrids from the 260 kb deletion of patient 3214 showed that all but one (exon 1) of the APC exons are removed by this deletion. PCR analysis of the somatic cell hybrid HHW1291, carrying the chromosome 5 homolog with the 100 kb deletion from patient 3824, revealed that exons 1 through 9 are present but exons 10

31

through 15 are missing. This result placed the deletion breakpoint either between exons 9 and 10 or within exon 10.

EXAMPLE 14

This example demonstrates the expression of alternately spliced APC messenger in normal tissues and in cancer cell lines.

Tissues that express the APC gene were identified by PCR amplification of cDNA made to mRNA with primers located within adjacent APC exons. In addition, PCR primers that flank the alternatively spliced exon 9 were chosen so that the expression pattern of both splice forms could be assessed. All tissue types tested (brain, lung, aorta, spleen, heart, kidney, liver, stomach, placenta, and colonic mucosa) and cultured cell lines (lymphoblasts, HL60, and choriocarcinoma) expressed both splice forms of the APC gene. We note, however, that expression by lymphocytes normally residing in some tissues, including colon, prevents unequivocal assessment of expression. The large mRNA, containing the complete exon 9 rather than only exon 9A, appears to be the more abundant message.

Northern analysis of poly(A)-selected RNA from lymphoblasts revealed a single band of approximately 10 kb, consistent with the size of the sequenced cDNA.

EXAMPLE 15

25

This example discusses structural features of the APC protein predicted from the sequence.

The cDNA consensus sequence of APC predicts that the longer, more abundant form of the message codes for a 2842 or 2844 amino acid peptide with a mass of 311.8 kd. This predicted APC peptide was compared with the current data bases of protein and DNA sequences using both Intelligenetics and GCG software packages. No genes with a high degree of amino arid sequence similarity were found. Although many short (approximately 20 amino arid) regions of sequence similarity were uncovered, none was sufficiently strong to reveal which, if any, might represent functional homology. Interestingly, multiple similarities to myosins and keratins did appear. The APC gene also was scanned for sequence motifs of known function; although

multiple glycosylation, phosphorylation, and myristoylation sites were seen, their significance is uncertain.

Analysis of the APC peptide sequence did identify features important in considering potential protein structure. Hydropathy plots (Kyte and Doolittle, J. Mol. Biol. Vol. 157, pp. 105-132 (1982)) indicate that the APC protein is notably hydrophilic. No hydrophobic domains suggesting a signal 10 peptide or a membrane-spanning domain were found. Analysis of the first 1000 residues indicates that α -helical rods may form (Cohen and Parry, Trends Biochem, Sci. Vol. 77. pp. 245-248 (1986); there is a scarcity of proline residues and, there are a number of regions containing heptad repeats (apolar-X-X-apolar-X-X-X). Interestingly, in exon 9A, the deleted form of exon 9, two heptad repeat regions are reconnected in the proper heptad repeat frame, deleting the intervening peptide region. After the first 1000 20 residues, the high proline content of the remainder of the peptide suggests a compact rather than a rod-like structure.

The most prominent feature of the second 1000 residues is a 20 amino acid repeat that is iterated seven times with semiregular spacing (Table 4). The intervening sequences between the seven repeat regions contained 114, 116, 151, 205, 107, and 58 amino acids, respectively. Finally, residues 2200–24000 contain a 200 amino acid basic domain.

30

TABLE IV

	Consensus:	F*VE*TP*CFSR*SSLSSLS
	1262:	YCVEDTPICFSRCSSLSSLS
35	1376:	HTVQETPLMFSRCTSVSSLD
	1492:	FATESTPDGFSCSSSLSALS
	1643:	YCVEGTPINFSTATSLSDLT
	1848:	TPIEGTPYCFSRNDSLSSLD
	1953:	FAIENTPVCPSHNSSLSSLS
	2013:	RHVEDTPVCFSRNSSLSSLS

Numbers denote the first amino acid of each repeat. The consensus sequence at the top reflects a majority amino acid at a given position.

5,137,806 8/1992 LeMaistie et al. . 5,244,801 9/1993 Tobi .

FOREIGN PATENT DOCUMENTS

WO 89/01481 8/1988 WIPO .

OTHER PUBLICATIONS

Kinzler et al, Science 253: 661-665, 1991.

Groden, et al., "Identification and Characterization of the Familial Adenomatous Polyposis Coli Gene", *Cell*, 66:589-600 (1991).

Joslyn, et al., "Identification of Delection Mutations and Three New Genes at the Familial Polyposis Locus", *Cell*, 66:601-613 (1991).

Kinzler, et al., "Identification of FAP Locus Genes From Chromsome 5q21", Science, 253:661-665 (1991).

Nishisho, et al., "Mutations of Chromosome 5q21 Genes in FAP and Colorectal Cancer Patients", Science, 253:665-669 (1991).

Orita, et al., Genomics, vol. 5, pp. 874-879, 1989.

Stanbridge, et al., "Identifying Tumor Suppressor Genes in Human Colorectal Cancer", Science, 247:12-13 (1990).

Fearon et al., "Identification of a Chromosome 18q Gene That is Altered in Colorectal Cancer", Science, 247:49-56 (1990).

Baker et al., "Chromosome 17 Deletions and p53 Gene Mutations in Colorectal Carcinomas", Science, 244:217-221 (1989).

Bodmer et al., "Localization of the Gene for Familial Adenomatous Polyposis on Chromosome 5", Nature, 328:614-616 (1987).

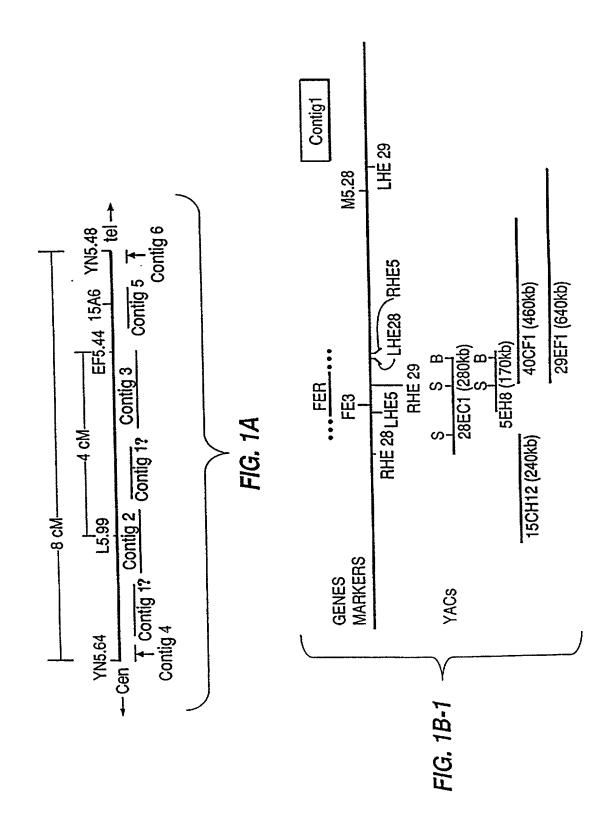
Primary Examiner—Lila Feisee
Assistant Examiner—Nancy A. Johnson
Attorney, Agent, or Firm—Banner & Witcoff, Ltd.

7] ABSTRACT

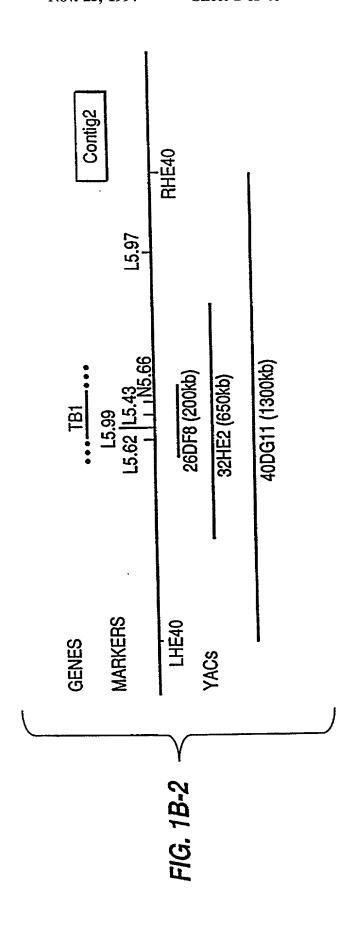
A human gene termed APC is disclosed. Methods and kits are provided for assessing mutations of the APC gene in human tissues and body samples. APC mutations are found in familial adenomatous polyposis patients as well as in sporadic colorectal cancer patients. APC is expressed in most normal tissues. These results suggest that APC is a tumor suppressor.

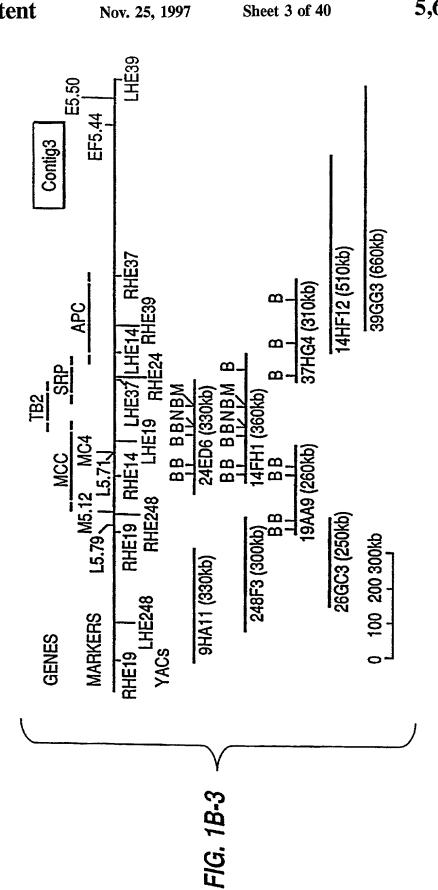
8 Claims, 40 Drawing Sheets

Nov. 25, 1997









F16. 24

U.S. Patent

TB1 AMINO ACID SEQUENCE

VAPVVVGSGR	VAPVVVGSGR APRHPAPAAM	HPRRPDGFDG	LGYRGGARDE	HPRRPDGFDG LGYRGGARDE QGFGGAFPAR SFSTGSDLGH	SFSTGSDLGH	09
WYTTPPDIPG	WYTTPPDIPG SRNLHWGEKS	PPYGVPTTST	PYEGPTEEPF	PPYGVPTTST PYEGPTEEPF SSGGGGSVOG OSSEOLNRFA	OSSECLNRFA	120
GFGIGLASLF	GFGIGLASLF TENVLAHPCI	VLRROCOVNY	HAOHYHLTPF	VLRROCOVNY HACHYHLTPF TVINIMYSFN KTOGPRALWK	KTOGPRALWK	180
GMGSTFIVOG	GMGSTFIVOG VTLGAEGIIS	EFTPLPREVL	HKWSPKOIGE	HKWSPKQIGE HLLLKSLTYV VAMPFYSASL	VAMPFYSASL	240
IETVOSEIIR	IETVOSEIIR DNTGILECVK	EGIGRVIGMG	VPHSKRLLPL	VPHSKRLLPL LSLIFPTVLH GVLHYIISSV	GVLHYIISSV	300
IOKFVLLILK	IOKFVLLILK RKTYNSHLAE	STSPVQSMLD	AYFPELIANF	STSPVOSMLD AYFPELIANF AASLCSDVIL YPLETVLHRL	YPLETVLHRL	360
HIGGIRTIID NTOLGYEVLP	NTOLGYEVLP	INTOYEGMRD CINTIRQEEG VFGFYKGFGA VIIOYTLHAA	CINTIROEEG	VFGFYKGFGA	VIIQY TLHAA	420
VLOITKIIYS TLLO	דרנס					434

F16. 2B

TB2 Amino Acid Sequence

120	180	185
ILKCGFLLW	KKATVNLLGE	
IFLSWFPFYY	ETADAITKEA	
VFSIAEFFSD	VVKDLKDKAK	
OWLTYWVYG	FLKHESOMDS	
IESPNKEDDT	LLYKRIIRPF	
GYPAYISIKA	MAPSPSNGAE	EKKST
	GYPAYISIKA IESPNKEDDT OMLTYWVYG VFSIAEFFSD IFLSWFPFYY ILKCGFLLWC 120	GYPAYISIKA IESPNKEDDT OWLTYWVYG VFSIAEFFSD IFLSWFPFYY ILKCGFLLWC 120 MAPSPSNGAE LLYKRIIRPF FLKHESOMDS VYKDLKDKAK ETADAITKEA KKATVNLLGE 180

Nov. 25, 1997

FIG. 3A

Leu Asn Ser Asn Met Lys Glu Val Leu 45 Gly TyrTyr Asp Gln Leu Leu Lys Gln Val Glu Ala 10 Ser Ser Ser Gly Glu Cys Ser 110 Glu Leu Asn Leu Asp Ser 75 Glu Leu Glu Asp Asn 30 Ser Ser Leu Arg Ala Glu Asp Glu Ala Met 60 Met Arg Lys 90 Ile Asp Leu Leu Glu Arg Leu Lys 70 Ser Asn Leu Arg Gln 25 Leu Thr Lys Leu Glu Thr Glu Ala 35 Ser 105 Lys Leu Arg Ser Gly Ser Val Ser Lys Gln Leu Gln Gly Ser Ile 50 Ser 5 Val 85 Glu Asn 20 G1yGlu 100 Met Ala Ala Ala Ser Arg Lys Met Asn

FIG. 3B

Ser Arg Glu GluGln Glu Glu Lys Glu Lys Asp Trp Tyr Tyr . 155 Thr 175 TyrSer Leu Cys Gln Gln Gly Ser Leu Pro Leu Glu 190 \mathtt{Thr} Ile Glu Leu Glu Lys Glu Arg 140 Pro Arg Arg Gly Phe Val Asn 120 Arg Arg Gln Leu Arg 220 Glu Glu Gln Leu Arg Arg Ile Ala Thr Lys Arg Ile Asp 170 Gln Thr Asp Leu Thr 185 Met 200 Ala Gly Tyr Leu Glu 135 Gln 215 Ala Asp Leu Asp Lys 150 Phe Ile Arg Val Glu Lys Arg Ala Ser Leu 165 Leu 180 G1yLeu Gln Asn Thr Met 115 Ser Pro Arg Leu 145 Asn Gln Ala Val Glu

FIG. 3C

Thr 240	Asp	Ala	Thr	Leu	Ser 320	Ala
Glu Lys Asp Ile Leu Arg Ile Arg Gln Leu Leu Gln Ser Gln Ala Thr 225	His 255	n Asn Glu Gly Gln Gly Val Gly Glu Ile Asn Met Ala 265	Glu	Thr His Ser Ala Pro Arg Arg Leu 300		
Gln	Ser	Asn 270	His	Arg	u Gly Thr Lys Val Glu Met Val Tyr Ser Leu Leu 310	His Asp Lys Asp Asp Met Ser Arg Thr Leu Leu 325
Ser	Ser Gln Asn Lys His Glu Thr Gly Ser 250	Ile	Asn Gly Gln Gly Ser Thr Thr Arg Met Asp His 280	Pro	Ser	Thr
Gln	Thr	Glu	Met	Ala 300	Tyr	Arg
Leu 235	Glu	Gly	Arg	Ser	Val 315	Ser
Leu	His 250	Val	Thr	His	Met	Met 330
Gln	Lys	G1y 265	Thr	Thr	G1 u	Asp
Arg	Asn	Gln	Ser 280	Ser	Val	Asp
Ile	Gln	Gly	G1y	Ser 295	Lys	Lys
Arg 230	Ser	G1u	Gln	Ser	Thr	Asp
Leu	Ser 245	Asn	Gly	Ser	Gly	His 325
Ile	Arg	G1 26	Asn	Leu	Leu	Thr
Asp	Glu	Arg	G1y 275	Val	His	Gly
Lys	Ala	G1u	Ser	Ser 290	Ser	Leu
Glu 225	Glu	Ala	Thr	Ala	Thr 305	Met

Cys	Val	Ser	G1y 400	Туг	Asp	Pro
Gly Cys	Ser	Ala	Arg	Ala 415	Met	Cys
Ser 350	Asp	Arg	Lys	Arg	G1y 430	Ile
Ile Ser Met Arg Gln 345	His Gly Asn Asp Lys Asp Ser 365	Asn Ser Arg Gly Ser Lys Glu Ala Arg Ala Arg Ala 375	Ile His Ser Gln Pro Asp Asp Lys Arg Gly 395	Val Leu His Leu Leu Glu Gln Ile Arg Ala Tyr 410	Cys Trp Glu Trp Gln Glu Ala His Glu Pro Gly Met Asp 420	Pro Val Glu His Gln Ile Cys
Arg	Asp	Arg 380	Asp	Gln	Glu	His
Met	Asn	Ala	Pro 395	Glu	His	Glu
Ser	Gly	Glu	Gln	Leu 410	Ala	Val
11e 345	His	Lys	Ser	Leu	Glu 425	Pro
Ser Gln Asp Ser Cys 340	Leu 360	Ser	His	His	Gln	Pro Ala 440
Ser	Leu Ile Gln Leu	G1y 375		Leu	Trp	Pro
Asp	Gln	Arg	11e 390	Val	Glu	sn Pro Met
Gln	Ile	Ser	is Asn	lle Arg 405	Trp	Pro
			His	Ile		Asn
Ser	Leu 355	61y	Leu	G1 u	Thr	Lys 435
Ser	Pro	Leu 370	Ala	Arg Arg	Glu	Gln Asp Lys 435
Met	Leu	Leu	Ala 385	Arg	Cys	Gln

FIG. 3E

His Glu His Arg Leu Ser Phe Asp Glu 460 Cys Val Leu Met Lys 455 Val 450

Gln 480 Glu Leu Leu Ala Ala Met Asn Glu Leu Gly Gly Leu Gln Ala Ile 465

 Thr Ser Tyr Gly Leu Thr Asn Asp His Tyr 490 Val Asp Cys Glu Met 485

Asp Met Arg Phe 510 Ser Met Lys Gly Cys 525 Thr Asn Leu Thr Leu 505 Lys Ala Thr Leu Cys 520 Leu Arg Arg Tyr 500 Asn 515 Ala

Ala Gly Met Ala

Gln Gln Val Leu 540 Glu Ser Glu Asp Ser 535 Ala Gln Leu Lys Val 530 Leu

Asp Val Asn Ser Ser

Val Leu Arg Asn Leu Ser Trp Arg Ala 550

FIG. 3F

Ala	Leu	Ala	Ser	Arg 640	Leu	His
Glu Val Gly Ser Val Lys Ala Leu Met Glu Cys Ala 565	Ala Leu	Ala His Cys Thr Glu Asn Lys Ala Asp Ile Cys Ala 600	Leu Ala Phe Leu Val Gly Thr Leu Thr Tyr Arg Ser 615	Leu Ala Ile Ile Glu Ser Gly Gly Gly Ile Leu Arg 630	Leu Ile Ala Thr Asn Glu Asp His Arg Gln Ile Leu 645	Ser
Glu	Ser 590	Ile	Tyr	Ile	Gln	Lys 670
Met	Lys Glu Ser Thr Leu Lys Ser Val Leu 585	Asp 605	\mathtt{Thr}	Gly	Arg	Cys Leu Gln Thr Leu Leu Gln His Leu Lys Ser 665
Leu	Val	Ala	Leu 620	Gly	His	His
Ala	Ser	Lys	Thr	G1y 635	Asp	Gln
Lys 570	Lys	Asn	Gly	Ser	Glu 650	Leu
Val	Leu 585	Glu	Val	Glu	Asn	Leu 665
Ser	Thr	${ m Thr}$	Leu	Ile	Thr	Thr
Gly	Ser	Cys	Phe 615	Ile	Ala	Gln
Val	Glu	His	Ala	Ala 630	Ile	ren
G1u 565	Lys	Ala	Leu	Leu	Leu 645	Cys
Arg	Lys 580	Ser	Gly Ala	Thr	Ser	Asn
Thr Leu Arg	Val	Leu 595		Asn	Ser	Asn
Thr	Glu ,	Trp Asn	Asp 610	Thr	Val	Arg Glu Asn
Lys	Leu	Trp	Val	G1n 625	Asn	Arg

FIG. 3G

Lys Asp Met Gly Ala 700 Ala Ser Gln Asn Leu Pro Lys Ala Ala Asn Arg Pro Ile Ala Trp 685 Lys Met Leu Cys Gly Thr Leu Ser Glu Leu Ile Asp Asn Leu Ser 780 Trp His 715 Ser G1yGlu Ala Leu Lys Leu Arg Asn Leu Met 730 Glu Ala Ser Ala 680 Ser His Leu 760 Asn Leu Ile 710 Gln 695 Lys Gln Lys Ala Phe Asp Asn 775 Ile Met Ser Pro Lys Asp Ala 725 Asn Leu Lys Asn Ile Val Tyr Lys Asp Ala 740 Gly Ser Ala Ala Thr Asn Thr 675 Arg 755 Glu Met Len Val Len His

FIG. 3H

Tyr Gly Asp Tyr Val 800 Thr Asn 815 Leu Ser Leu Glu Arg Glu Arg Gly Ile Gly Leu Gly Asn Tyr 855 Arg Gly Leu Gln Ser Asp Asn Phe Val 830 Ser Glu Val \mathtt{Thr} Arg 845 Thr Ser Glu Leu 795 Leu Asn Ser Arg Gly Ser Leu Asp Ser 840 Met Lys Gln Ser His Asp Asp Asn Arg 810 Ser Val 890 Val Leu Ser Pro Tyr 825 Ser Lys Gly Thr Ile Ala His Gln Arg 790 Gln Thr Asn Arg 805 Ala 885 Thr Glu Asn LysThr 820 Ser Thr Ala Ser Gly Asn Met Ser 835 Phe Asp Arg Ser Ala ThrSer

FIG. 31

Leu	Ala	Asn	Ser 960	Arg	Ser	Ile
Ser Gln Glu Asp Arg Ser Ser Gly Ser Thr Thr Glu Leu 900	Ser Ala Ala	Ser		Ser Val Ser Ser Asn Asp Gly Tyr Gly Lys Arg 970	Glu	Ser Tyr Gly Gln Tyr Pro Ala Asp Leu Ala His Lys 1000
$_{910}$	Ser	Asn	Arg	Gly	Asp 990	His
Thr	Ser 925	Glu Asn Ser	Pro Tyr Ala Lys Leu Glu Tyr Lys Arg Ser 950	Tyr	Lys Pro Ser Ile Glu Ser Tyr Ser Glu Asp Asp Glu 980	Ala 1005
Ser	Arg	Ser 940	$\mathbf{T}\mathbf{y}\mathbf{r}$	Gly	Glu	Leu
Gly	Arg	Lys	G1u 955	Asp	Ser	Asp
Ser	Thr Asp Glu Arg Asn Ala Leu Arg Arg Ser 920	Ser Asn Thr Tyr Asn Phe Thr Lys Ser 935	Leu	Asn 970	Tyr	Ala
Ser 905	Ala	Phe	Lys	Ser	Ser 985	Pro
Arg	Asn 920	Asn	Ala	Ser	Glu	Tyr 1000
Asp	Arg	Tyr 935	Tyr	Val	Ile	Gln
Glu	Glu	Thr	Pro 950	Ser	Ser	Gly
Gln	Asp	Asn	Ser Met	Asn 965	Pro	Tyr
Ser 900	Thr	Ser	Ser	Leu	Lys 980	Ser
Thr	Val 915	His	Cys	Ser	Met	Cys 995
Ile His	Cys	Thr 930	Thr	Asn Asp	Gly Gln Met	Phe
Ile	His	His	Arg 945	Asn	Gly	Lys

FIG. 3J

Ser Ala Asn His Met Asp Asp Asn Asp Gly Glu Leu Asp Thr 1010 His

Ile Asn Tyr Ser Leu Lys Tyr Ser Asp Glu Gln Leu Asn Ser Gly Arg 1025 1040

Nov. 25, 1997

Gln Ser Pro Ser Gln Asn Glu Arg Trp Ala Arg Pro Lys His Ile 1055

Glu Asp Glu Ile Lys Gln Ser Glu Gln Arg Gln Ser Arg Asn Gln Ser 1060

Tyr Pro Val Tyr Thr Glu Ser Thr Asp Asp Lys His Leu Lys 1075 Thr ${
m Thr}$

Gln Pro His Phe Gly Gln Gln Glu Cys Val Ser Pro Tyr Arg Ser 1090 Phe

Arg Gly Ala Asn Gly Ser Glu Thr Asn Arg Val Gly Ser Asn His 1105

FIG. 3K

Glu	Gln	Glu	Ala	Ser Ser Gln Lys Gln Ser Phe Ser Phe Ser Lys Ser 1190	Glu	His
Ile Asn Gln Asn Val Ser Gln Ser Leu Cys Gln Glu Asp Asp Tyr Glu 1125	Pro Thr Asn Tyr Ser Glu Arg Tyr Ser Glu Glu Glu Gln 1140	Glu Glu Glu Arg Pro Thr Asn Tyr Ser Ile Lys Tyr Asn Glu 1155	Val Asp Gln Pro Ile Asp Tyr Ser Leu Lys Tyr Ala 1175	Lys	Ser Ser Lys Thr Glu His Met Ser Ser Ser Glu 1205	Thr Pro Ser Ser Asn Ala Lys Arg Gln Asn Gln Leu His 1220
Asp	Glu 115($^{\mathrm{Tyr}}_{5}$	Lys	Ser	Ser	Gln 1230
Asp	G1u	Lys 1169	Leu	Phe	Ser	Asn
Glu	Ser	Ile	Ser 118(Ser	Ser	Gln
Gln 0	Tyr	Ser	Tyr	Phe 1195	Met	Arg
Cys 113	Arg 5	Tyr	Asp	Ser	His 121(Lys
Leu	Glu 114	Asn o	Ile	Gln	Glu	Ala 1225
Ser	Ser	Thr 116	Pro	Lys	Thr	Asn
Gln	${ m T} { m V} { m r}$	Pro	Gln 117	Gln	Lys	Ser
Ser	Asn	Arg	Asp	Ser 1190	Ser	Ser
Val 1129	$\operatorname{\mathtt{Thr}}_{\mathfrak{I}}$	G1 u	Val	Ser	Ser 1205	Pro)
Asn	Pro 114	Glu 5	His	Pro	Gln	Thr 122(
Gln	Lys	G1u 1159	Glu Lys Arg His 1170	11e	Gly Gln	Ser
Asn	Asp Asp Lys	Glu	Lys 1170	Thr Asp 1185	Ser	Thr
Ile	Asp	His	Glu	Thr / 1185	Ser	Asn

FIG. 3L

Glu Asp Thr Pro Ile Cys Phe Ser Arg Cys Ser Ser Leu Ser Ser Leu Thr Ile Asn Gln Glu Thr Ile Gln Thr Tyr Cys Val Ser Ala Glu Asp Glu Ile Gly Cys Asn Gln Thr Thr Gln Glu Ala Asp Ser Ala Asn Thr Leu Gln Ile Ala Glu Ile Lys Gly Lys Ile Gly Thr Arg Ser Ala Glu Asp Pro Val Ser Glu Val Pro Ala Val Ser Gln Ser Ala Gln Ser Arg Ser Gly Gln Pro Gln Lys Ala Ala Pro Arg Thr Lys Ser Ser Arg Leu Gln Gly Ser Ser Leu Ser Lys Val Ser Ser Pro Ser Ser His

FIG. 3M

Asp Ser Pro Gly Gln Thr Met Pro Pro Ser Arg Ser Lys Thr Pro Pro 1425 Ser Gly Ala Gln Thr Pro Lys Ser Pro Glu His Tyr Leu Asp Ser Phe Glu Ser Arg Ser Ile Ala Ser Ser Val Gln Ser Glu Pro Pro Pro Gln Thr Ala Gln Thr Lys Arg Glu Val Pro Lys Asn Lys Ser His Lys Ala Val Glu Phe Pro Ser Gly Ala Lys Ser Gly Ile Ile Ser Pro Ser Asp Leu Val Gln Glu Thr Pro Leu Met Phe Ser Arg Cys Thr Ser Val (Pro Cys Ser Gly Met Val Glu Ser Ala Arg Pro Ser Lys

FIG. 3N

dentification of the second of

Asn Ala Ala Val Gln Arg Val Gln Val Leu Pro Asp Ala Asp Thr Leu 1475 Ser Glu Ser Leu Ser Ala Leu Ser Leu Asp Glu Pro Phe Ile Gln Lys Asp Val 1505 Ala Glu Lys Thr Ile Asp Ser Glu Lys Asp Leu Leu Asp Asp Ser Asp 1555 Glu Leu Arg Ile Met Pro Pro Val Gln Glu Asn Asp Asn Gly Asn (1535) Pro Thr Ala Glu Lys Arg Glu Ser Gly Pro Lys Gln Ala Ala 1460 Thr Glu Ser Glu Gln Pro Lys Glu Ser Asn Glu Asn Glu Lys 1540 Leu His Phe Ala Thr Glu Ser Thr Pro Asp Gly Phe Ser Cys Ser 1490 1495 1490 Ala

FIG. 30

Thr Lys Ser Ser Arg Lys Gly Lys Lys Pro Ala Gln Thr Ala Ser Lys Leu Pro Pro Val Ala Arg Lys Pro Ser Gln Leu Pro Val Tyr Lys Thr Pro Gly Asp Asp Met Pro Arg Val Tyr Cys Val Glu Gly Thr Pro Asp Asp Ile Glu Ile Leu Glu Glu Cys Ile Ile Ser Ala Met Pro Leu Leu Pro Ser Gln Asn Arg Leu Gln Pro Gln Lys His Val Ser Phe Ile Asn Phe Ser Thr Ala Thr Ser Leu Ser Asp Leu Thr Ile Glu Ser Pro Pro Asn Glu Leu Ala Ala Gly Glu Gly Val Arg Gly Gly Ala Gln

FIG. 3P

Lys Ile Met Asp Gln Val Gln Gln Ala Ser Ala Ser Ser Ala Pro Ser Gly Glu Phe Glu Lys Arg Asp Thr Ile Pro Thr Glu Gly Arg Ser Thr Asp Glu Ala Gln Gly Gly Lys Thr Ser Ser Val Thr Ile Pro Glu Leu Asp Asp Asn Lys Ala Glu Glu Gly Asp Ile Leu Ala Glu Cys Ile Asn Ser Ala Met Pro Lys Gly Lys Ser His Lys Pro Phe Arg Val Lys Asn Lys Asn Gln Leu Asp Gly Lys Lys Lys Lys Pro Thr Ser Pro Val Lys Pro Ile Pro Gln Asn Thr Glu Tyr Arg Thr Arg Val Arg Lys Asn 1780

1890

FIG. 3Q

Asn Lys Asp Ser Lys Lys Gln Asn Leu Lys Asn Asn Ser Lys Asp Phe Asn 1810 Asp Ser Pro His His Tyr Thr Pro Ile Glu Gly Thr Pro Tyr Cys Phe 1845 Gln Ser Arg Asn Asp Ser Leu Ser Ser Leu Asp Phe Asp Asp Asp Val 1860 Asp Leu Ser Arg Glu Lys Ala Glu Leu Arg Lys Ala Lys Glu Asn Lys 1875 Ala Asp Ser Lys Asn Asn Leu Asn Ala Glu Arg Val Phe Ser Asp 1795 Asp Lys Leu Pro Asn Asn Glu Asp Arg Val Arg Gly Ser Phe Ala 1825 Glu Leu Thr Ser Asn Thr Ser His Thr Ser Glu Ala Lys Val Glu

FIG. 3R

Gln Ser Ala Asn Lys Thr Gln Ala Ile Ala Lys Gln Pro Ile Asn Arg Pro Gln Ser Ser Lys Asp Ile Pro Asp Arg Gly Ala Ala Thr Asp Glu Lys Leu Gln Ser Leu Ser Ser Leu Ser Asp Ile Asp Gln Glu Asn Asn Lys Glu Asn Glu Pro Ile Lys Glu Thr Glu Pro Pro Asp Ser Gln Gly Glu Pro Ser Phe Ala Ile Glu Asn Thr Pro Val Cys Phe Ser His Asn Ser Lys Pro Gln Ala Ser Gly Tyr Ala Pro Lys Ser Phe His Val Gly Gln Pro Lys Pro Ile Leu Gln Lys Gln Ser Thr Phe 1945 Asn

FIG. 3S

Pro Arg Asn Met Gly Gly Ile Leu Gly Glu Asp Leu Thr Leu Asp Leu 2065 Ser Ala Met Pro Ser Lys Asp Ile Gln Arg Pro Asp Ser Glu His Gly Leu Ser Pro Asp Ser 2095 Ser Ser Leu His Gln Ala Ala Ala Ala Ala Cys Leu Ser Arg Gln Ala 2115 Glu Asn Phe Asp Trp Lys Ala Ile Gln Glu Gly Ala Asn Ser Ile Val 2100 Leu Ser 2030 Lys Lys Lys Pro Ser Arg Leu Lys Gly Asp Asn Glu Lys His 2050 2045 Ser Asp Ser Glu Asp Asp Leu Leu Gln Glu Cys Ile Ser 2035 Pro Val Cys Phe Ser Arg Asn Ser Ser Leu Ser 2020 Thr

FIG. 3T

Leu	Phe His Leu Thr Pro Asp Gln Glu Glu Lys Pro Phe Thr 2150	Leu	Lys	Glu	Ile	Arg Gly Arg Thr Met Ile His Ile Pro Gly Val Arg Asn Ser Ser 2230 2240
Ser	Phe	Thr 2175	Gly	Ser	Ser	Ser
Ser Asp Ser Ile Leu Ser Leu Lys Ser Gly Ile Ser Leu 2135	Pro	Gly Pro Arg Ile Leu Lys Pro Gly Glu Lys Ser Thr Leu 2165	Lys Ile Glu Ser Glu Ser Lys Gly Ile Lys Gly Gly Lys 2180	Lys Ser Leu Ile Thr Gly Lys Val Arg Ser Asn Ser Glu 2200	Gln Met Lys Gln Pro Leu Gln Ala Asn Met Pro Ser Ile 2215	Asn
Gly	Lys	Lys	Lys	Ser 2205	Met	Arg
Ser 214(Glu	Glu	Ile	Arg	Asn 2220	Val
Lys	Glu 2155	Gly)	Gly	Val	Ala	G1y 2235
Leu	Gln	Pro 2170	Lys	Lys	Gln	Pro
Ser	Asp	Lys	Ser 2185	Gly	Leu	Ile
Leu	Pro	Leu	Glu	Thr 2200	Pro	His
11e 2139	Thr)	Ile	Ser	Ile	Gln 2215	Ile
Ser	Leu 215(Arg	Glu	Leu	Lys	Met 2230
Asp	His	Pro 2165	Ile	Ser	Met	Thr
Ser	Phe	Gly	Lys 2180	Lys	Gln	Arg
Ser Asp 2130	Pro	Lys	Lys	Tyr 2195	Glγ	Glγ
	Ser	Asn	Thr Lys	Val	Ser 2210	Arg
Ser	Gly 2	Ser	Glu	Lys	Ile	Ser 1 2225

Sheet 26 of 40

FIG. 3U

FIG. 3V

Glu Ser Ala Ser Lys Gly 2395 Ser Lys Gln Thr Gly Leu 2380 Leu Asn Gln Met Asn Asn Gly Asn Gly Ala Asn Lys Lys Val Glu Leu 2405 Ser Arg Met Ser Ser Thr Lys Ser Ser Gly Ser Glu Ser Asp Arg Ser 2420 Ser \mathtt{Thr} Glu Arg Pro Val Leu Val Arg Gln Ser Thr Phe Ile Lys Glu Ala 2435 Glu Gly Ser Gly Lys Met Ser Tyr 2360 Phe Ala Ser 2460 Thr Pro Thr Leu Arg Arg Lys Leu Glu Glu Ser 2450 Ser Gln Gln Asn Leu Ser Ile Pro Arg Ser 2390 2375 Ser Ser Lys Ser Gly Arg Gln Met 2370 Ser Thr 2355 Ser Lys Asn Ala Ala Pro Ser

FIG. 3W

Leu Ser Pro Ser Arg Pro Ala Ser Pro Thr Arg Ser Gln Ala Gln 2465 Gly Thr Trp Lys Arg Glu His Ser Lys His Ser Ser Leu Pro Arg Thr Pro Val Leu Ser Pro Ser Leu Pro Asp Met Ser Leu Ser Thr His Ser Val Gln Ala Gly Gly Trp Arg Lys Leu Pro Pro Asn Leu Ser Ile Glu Tyr Asn Asp Gly Arg Pro Ala Lys Arg His Asp Ile Arg Ser His Ser Glu Ser Pro Ser Arg Leu Pro Ile Asn Arg Ser Val Ser Thr Trp Arg Arg Thr Gly Ser Ser Ser Ser Ile Leu Ser Ala Pro Thr Ser

FIG. 3X

Glu Ser Ser Glu Lys Ala Lys Ser Glu Asp Glu Lys His 2580 2585 Ser Ser

Asn Ser Ile Ser Gly Thr Lys Gln Ser Lys Glu Asn Gln Val Ser Ala 2595

Lys Gly Thr Trp Arg Lys Ile Lys Glu Asn Glu Phe Ser Pro Thr Asn 2610 2620

Ser Thr Ser Gln Thr Val Ser Ser Gly Ala Thr Asn Gly Ala Glu Ser 2625 264

Lys Thr Leu Ile

Tyr Gln Met Ala Pro Ala Val Ser Lys Thr Glu Asp 2645 2655

Val Trp Val Arg Ile Glu Asp Cys Pro Ile Asn Asn Pro Arg Ser Gly 2665

Arg Ser Pro Thr Gly Asn Thr Pro Pro Val Ile Asp Ser Val Ser Glu 2675

FIG. 3Y

Ser Ala Asp Ser Thr Ser Ala Lys Ala Asn Pro Asn Ile Lys Asp Ser Lys Asp Asn Gln Ala Lys Gln Asn Val Gly Asn Gly Ser Val Pro Met Arg Thr Val Gly Leu Glu Asn 2705 Phe Ile Gln Val Asp Ala Pro Asp Gln Lys Gly Thr Glu Ile Lys Pro Gly Gln Asn Asn Pro Val Pro Val Ser Glu Thr Asn Lys His Ser Ser Pro Ser Gly Thr Val Ala Ala Arg Val Thr Pro Phe Asn Tyr Asn Pro Ser Pro Arg Lys Ser Arg Leu Thr Ser

FIG 37

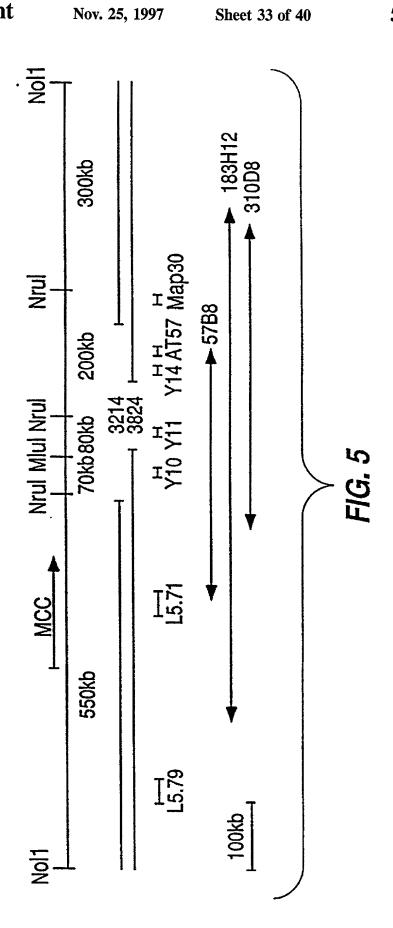
-- SI302

Asn Asn Thr Lys Lys Arg 2810	Gly Thr Gln Ser Pro Lys 2830	Val
Asn 281	Ser	Ser
Val	Ser 2825	
Pro Val	Glu	Val Thr 2840
Pro Thr	Thr	
Pro	Ser	Tyr Leu
Ile 2805	Asp)	Ser
Gln	Thr 2820	Gly
Ser	Lys	Ser 2835
Pro	Ser	His
Arg	Asp	Arg

Nov. 25, 1997

F16. 44

APC	203	LGTCODMEKRAORRIARIOOIEKDILRIROI	233
RAL2	576	LTGAKGLOLRALRRIARIEOGGTAISPTSPL	909
		F1G. 4B	
APC	453		481
н3 МАСНR	249		277
MCC	220	LYPNLAEERSRWEKELAGLREENESLTAM	248
APC	453		481



55 ACG 109 GAG GBU 163 AAC ASD 177 177 177 176 116 433 433 487 ATC 116 541 CTT CAC GAG GTG GTG Val GCC TGG GAT TTG ATC Ile GGC Gly CTG GGC Gly CTG CCA Pro CAG Gln TCT CTG Leu GGN TTC ACC Thr TAC ACC TTC Phe TAC TTC AAG Lys GTC Val CCG CGG AAA Lys TTG GGC Gly GAT GGC Gly ${
m TTC}$ TACGTG GCC GCC GAC GCC TTT Phe GAT Asp GAA Glu TGTCTC AGT TTC Phe ACA GTG Val GAG GGA Gly GAA Glu AAG Lys GCT CTG Leu GAC GGA Gly AGG CTC CTG Leu ATA Ile AAA Lys ATT Ile CTG Leu GAA ATG MET CTA GAG Glu AAG Lys GGA Gly CTG Leu AAC Asn AGC Ser ATG MET GCT CAG Gln AGG Arg 136 GCC Ala 190 ATC Tle AAC ASD 298 CCC Pro 352 TTC Phe 406 TAC 460 GGG Gly 514 TCC Ser ccg Pro CTG Leu ATG TGC GTC AGT Ser TAC GTG Val AAT TATGCC CTT GGT Gly CTC Leu GAG Glu rrc Phe GGT Gly TCT Ser GAC GCG CTT Leu CTC Leu ATA Ile CCC TAT Tyr AAG Lys CCT CCA Pro ACT TCT GCT TTC Phe TCT GCT GTG Val AGC CTG GCT ATG MET ATC Ile ATG MET GCC AAA Lys ${f TGG}$ CCG GTA Val GCC GCC TGC Cys TTC Phe GGA Gly ATT Ile TGG TCA GTC CCC AAC AGC TCA CTG Leu TAT TAC ATG GTC AAG Lys AGG GGT Gly ACC

F16. 6B

595					700	TTATATTAGG	770	TGGAATGTGT	840	CAGTGGGCAG	910	CTGCAGGAAA	086	CACGSATTTT	1050	ATAATTCNGR	1120	TGCATCATGC	1190	CACCTGCCAA	1260	TTAATATGCA	1330	GGCATATGAA	1400
	GCG AAG AAA	Ala Lys Lys	AGA		069	GAGCTTGATG	760	TATTAAAGAT	830	AAACTTAATG	006	TGTTGCTATC	970	GCTCTCCCTG	1040	ACAATTTTAT	1110	GACTACANCA	1180	ACAGTAAGAC	1250	AAATACGTGA	1320	CGTAGTATAT	1390
	ACT AAA GAA	Thr Lys Glu	ACC TAA ACC	\mathtt{Thr}	089	CTTCCTACTG	750	ATTTTTGAGA	820	GGAGCACTTT	890	AAAAGATGTA	096	ACTTTACTGG	1030	CCTRTAATGT	1100	GTTACTGTCT	1170	TAACTTCTGT	1240	ATACTTTAGG	1310	TGGTTGTTTC	1380
89	GAT GCC ATC	Ala Ile	AAG AGC	Lys Lys Ser		CTCTCTGTAC	740	CCTTGGAAAC	810	ATATATAG	880	TCTGGGTAGG	950	CAGGCTGTGT	1020	GGTTCTACTT	1090	ATATGGAAAT	1160	GTGTCATTTA	1230	CTACTAAATA	1300	GAGATTGGCC	1370
56	ACT GCA	Thr Ala	GAA GAA	Glu Glu	99	ACTTCCTGCC	730	AATAATGTTG	800	TTTACTGTCT	870	GTATTTGCC	940	ATATACCCCA	1010	TAATCTTTAT	1080	GCACATGTAC	1150	AAGGTTGTAT	1220	CTGGTGTGGT	1290	AAATCGAATG	1360
	TCC AAA GAG		TTA CTG GGT	Leu Leu Gly	650	CTGGATGGAA	720	TAATTATTT	790	TTTGCTTACT	860	TTTGGAAAAT	930	AAATAAAATT	1000	ACATTTAGGR	1070	ATGTATTTGT	1140	GGAGCA	1210	AACCATTGTG	1280	GTGAGAAATG	1350
	AAA GAC AAG	Lys Asp Lys	ACC GTG AAT	Thr Val Asn	640	CTAAACCAGA	710	GACTGTGGTA	780	TGTAAGTTTC	850	TGTCCACGTT	920	TATAAACTTA	066	CTCTGTAGTT	1060	AATGTTTTA	1130	TCATGGGGAG	1200	AAGCTGGAGG	1270	AGTGAACAAA	1340

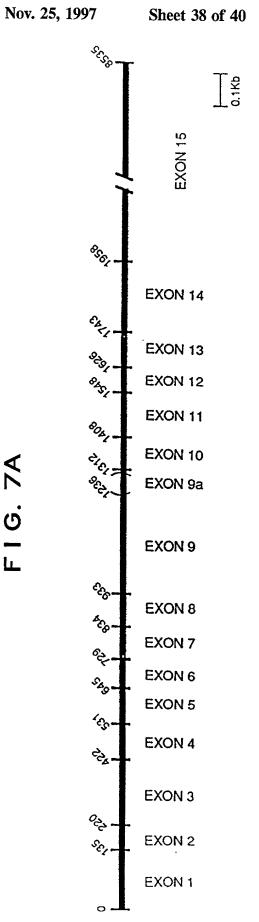
F/G. 6C

GCTTTATAAA GCAGTTAGTT
CACACACA CACACACACA
AACTAGTAAT ACTGTCTTAT
1560
CCATTTCTGG TTTTATCTTC
1630
AMCCAGTTTN AGGMNCTTCT
1700
CAACAACATG CTAATGRCGA
GCTTGGCAAT
CCTAGTTTAC
CTGCACAKGA
GNTATAGAGA
TTAAACTAGA
2130
AAGGCAAATA
TATA ANA ANT STATE STATE STATE OF STATE

5/6. 6D

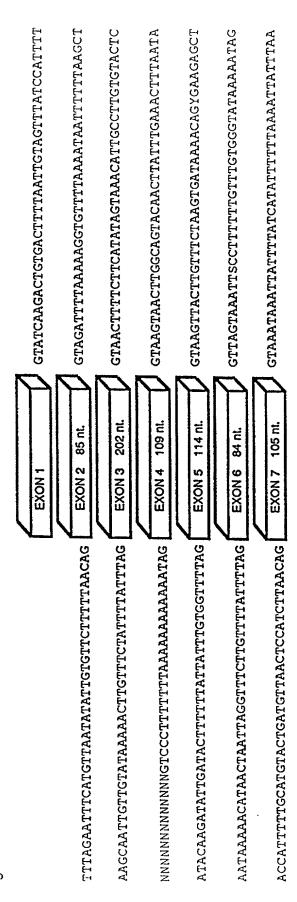
		•				
2250	2260	2270	2280	2290	2300	2310
Ē	AGAGCATT	TTTTGGGGAG	GAAAACAGTG	GTGATTACCG	GCATTTTATT	AAACTTAAAA
2320			2350	2360	2370	2380
CTTTGTAGAA	AGCAAA	ATTGTTC	GGAGAAAATC	AACTTTTAGA	TTAAAAAAT	AW
2390			2420	2430	2440	2450
TAGGAGTATT	TAAATC	TCCCATAAAT	AAAAGT	TTTTCTTGGT	GGCAGAATGA	A.
2460		2480		2500	2510	
CNTCTAGCAT	ATAGAC	TAATCAGATT	GACAGO	AGAATATATT	ATCAGACAAG	ATGAGGAG
2530		2550		2570	2580	
ACAAAAGTTA	CTATTGC	TAATGACTTA	CAGGCTAAAA	NTAGNTNTAA	AATACTATAT	TAAATT
2600		2620		2640		2660
ATGCAATTTT	TTTTTG	CTTGAGACCA	AAATTTAAGT	TAACTGTTGC	TGGCAGTCTA	AGTGTAAATG
2670		2690		2710		2730
TTAACAGCAG	GAGAAG	GAATTGAGCA		ATGATTTCCC	AAATGAAATA	CIGCCIIGGC
2740		2760		2780	2790	2800
TAGAGTTTGA	AAAACT	GAGCCTGTGC	CTGGCTAGAA	AACAAGCGTT	TATTTGAATG	TGAATAGTGT
2810		2830	2840	2850	2860	2870
TTCAAAGGTA	TGTAGI	GAATTCCTAC	CAAACAGCTT	AAATTCTTCA	AGAAAGAATT	CCTGCAGCAG
2880		2900	2910	2920	2930	2940
TTATTCCCTT	ACCTGAAGGC	TTCAATCATT	TGGATCAACA	ACTGCTACTC	TCGGGAAGAC	TCCTCTACTC
2950		2970	2980	2990	3000	3010
ACAGCTGAAG	AAAATG	CACCCTTCAC	ACTGTTATCA	CCTATCCTGA	AGATGTGATA	CACTGAATGG
3020	3030	3040	3050	3060	3070	α
AAATAAATAG	ATGTAAATAA	AATTGAGWTC	TCATTTAAAA	AAAACCATGT	GCCCAATGGG	Ü
3090	3100	3110	3120	3130	3140	312
CATGTTGTGG	TTTAAACAGC	AACTGCACCC	ACTAGCACAG	CCCATTGAGC	TANCCTATAT	ATACATCTCT
3160						
GTCAGTGCCC	CIC					





ñ

FIG. 7B-



-16. 7B-2

TAGTCTAAATTTTTAG	EXON 8 99 nt.	GTAACAGAAGATTAÇAAACCCTGGTCACTAATGCCATGAC
TAAAGTCGTAATTTTGTTTCTAAACTCATTTGGCCCACAG	EXON 9 379 nt.	GIATGTTCTCTATAGTGTACATCGTAGTGCATGTTTCAAA
ATAACAAAGCATTATGGTTTATGTTTTTTTTTCAG	EXON 10 96 nt.	GTAAGACAAAATGTTTTTAATGACATAGACAATTACTG
TIAGAIGATTGTCTTTTCCTCTTTGCCCTTTTTAAATTAG	EXON 11 140 nt.	GTATGTTTTTATAACATGTATTTCTTAAGATAGCTCAGGT
TGNCTTTTAAATGATCCTCTATTCTGTATTTAAATTTACAG	EXON 12 78 nt.	GTACTATTTAGAATTTCACCTGTTTTTTCTTTTTTTTTT
ATTTTATGTATAAATTAATCTAAAATTGATTAATTTCCAG	EXON 13 117 nt.	GTACCTTTGAAAACATTTAGTACTATAATATGAATTTCAT
CCAACTCNAATTAGATGACCCATATTCTGTTTCTTACTAG	EXON 14 215 nt.	GTATATATAGAGTTTTATATTACTTTTAAAGTACAGAATT
ATTGTGACCTTAAFTTTGTGATCTCTTGATTTTATTTCAG	EXON 15	

JOINT DECLARATION FOR REISSUE PATENT APPLICATION

a por

As the below named inventor, we hereby declare that:

Our residence, post office address and citizenship are as stated below next to our names;

We believe we			joint inventors	of the subject matter v	which is claimed and for	which a patent is sought on the
ne specification of which		JIES				
_	n ached hereto					
	filed on <u>Ma</u>		•	as Application Serial 1	Number <u>08/452,654</u>	and was amended o
		if applicable		us rippiiouson sonur	<u> </u>	and was amonded o
We hereby sta mended by any amenda			ed and understa	and the contents of the	above identified specific	ation, including the claims, a
We acknowle legulations, §1.56(a).	dge the duty	to disclose	information w	which is material to pate	entability in accordance	with Title 37, Code of Federa
			Prior Fo	oreign Application('s)	
We hereby claisertificate listed below a hat of the application or	ind have also	identified	below any fore	United States Code, §1 ign application(s) for p	19 of any foreign applicatent or inventor's certificatent	ation(s) for patent or inventor' cate having a filing date befor
Country		Applicat	ion Number	Date of Filing (day, month, year)	Date of Issue (day, month, year)	Priority Claimed Under 35 U.S.C. §119
United Kingdo	om	910	0962.1	16/01/91		YES
United Kingdo	om	910	0963.9	16/01/91		YES
United Kingdo	om	910	0974.6	16/01/91		YES
United Kingdo	om	910	0975.3	16/01/91		YES
ubject matter of each of aragraph of Title 35, Un	the claims of ited States Co	this applicated this thick the state of the	, United States ation is not disc e acknowledge	closed in the prior Unit the duty to disclose mate	ited States application(s) ed States application in the crial information as defin	listed below and, insofar as the manner provided by the first and in Title 37, Code of Federal international filing date of the
Application S	erial Number			Date of Filing (Day, Month, Year)		atus — Patenied, ding, Abandoned
We hereby claim player also identified below	any provisi	fits under T onal applica	itle 35, United	tes Provisional App States Code, §119(e) o having a filing date be te of Filing	f any provisional applica fore that of the applicatio	tion for patent listed below an n on which priority is claimed rity Claimed
TF				, month, year)		U.S.C. §119(e)

5,691,454, is wholly or partially inoperative or invalid because of the following defects in the specification:

- the amino acid sequence provided for the APC protein in SEQ ID NO:7 of the sequence listing contains a minor error; and
- the specification refers to overlapping APC cDNA clones as "defining an ORF of 2842 amino acids" (column 4, line 31) and as coding "for a 2842 or 2844 amino acid peptide" (column 31, lines 32-33), rather than the correct number of 2843 amino acids.
- (2) The correction of SEQ ID NO:7 is supported by the specification. The missing proline at position 173 in SEQ ID NO:7 is supported in the specification by the proline which is present at position 173 in SEQ ID NOS:1 and 2 and in Figure 3. In addition, routine analysis of YAC 37HG4 deposited as NCIMB 40353, referred to at column 12, lines 35-39 of U.S. Patent 5,691,454 establishes that there is, indeed, a proline at codon 173. The deposit was made under the terms of the Budapest Treaty. (See declaration of Dr. Sarah Kagan, of record in Serial No. 08/452,654, filed February 14, 1996.) One of ordinary skill in the art would have recognized the omission of the proline in SEQ ID NO:7 as a minor error by noting the inconsistency between the amino acid sequences presented in Figure 3 and in SEQ ID NOS:1 and 2 with that in SEQ ID NO:7.
- (3) The error at column 4, line 31, referring to "an ORF of 2842 amino acids," occurred because of the inadvertent omission of the proline at position 173 in originally filed Figure 3. The omission of this proline resulted in the APC protein being described in the specification as having 2842 rather than 2843 amino acids.
- (4) The error at column 31, lines 32-33, referring to a "2842 or 2844 amino acid peptide," occurred as follows. The application which issued as U.S. Patent 5,691,454 originally contained eight figures. In Figure 7 as originally filed, three supernumerary nucleotides were added at nucleotide positions 3972 (C), 3981 (G), and 3996 (A). As a result, the predicted amino acid sequence was erroneously stated to be "Ser Ser Val His Ser Thr Leu Glu" rather than "Ala Val Ser Gln His Pro Arg" at positions 1325 to 1331. This error resulted in an apparent sequence for the APC protein of 2844 amino acids. In combination with the omission of the proline at position 173 in originally filed Figure 3, this error resulted in the APC protein being described in the specification as a "2842 or 2844 amino acid peptide." Originally filed Figure 7 was canceled during prosecution of Serial No. 08/452,654, which issued as U.S. Patent 5,691,545.
- (5) Correction of the number of amino acids in the APC protein does not add new matter to the specification. It merely renders consistent the number of amino acids shown in SEQ ID NOS:1 and 2 and the number of amino acids referred to in the specification.
- (6) All errors which are being corrected in the present reissue application up to the time of filing of this declaration arose without any deceptive intent on the part of the applicants.
- (7) We hereby declare that all statements made herein of our own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application

And we hereby appoint, both jointly and severally, as our attorneys with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected herewith the following attorneys who are all members of the Bar of the District of Columbia, their registration numbers being listed after their names:

Donald W. Banner, Registration No. 17,037; Edward F. McKie, Jr.. Registration No. 17,335; William W. Beckett, Registration No. 18,262; Dale H. Hoscheit, Registration No. 19,090; Joseph M. Potenza, Registration No. 28,175; James A. Niegowski, Registration No. 28,331; Joseph M. Skerpon, Registration No. 29,864; Thomas L. Peterson, Registration No. 30,969; Nina L. Medlock, Registration No. 29,673; William J. Fisher, Registration No. 32,133; Thomas H. Jackson, Registration No. 29,808; Patricia E. Hong, Registration No. 34,373; Robert S. Katz, Registration No. 36,402, Brian E. Hanlon, Registration No. 40,449, Sarah A. Kagan, Registration No. 32,141 and Lisa M. Hemmendinger, Registration No. 42,653.

All correspondence and telephone communications should be addressed to: Banner & Witcoff, Ltd., Eleventh Floor, 1001 G Street, N.W., Washington, D.C. 20001-4597, telephone number (202) 508-9100, which is also the address and telephone number of each of the above listed attorneys.

Signature	reill		Date h 30 97	_
Full Name of				
First Inventor	WOGELSTEIN Family Name	Bert First Given Name	Second Given Name	_
Residence	3700 Breton Way, Baltimore,	Maryland 21208		
Citizenship Post Office	United States of America			_
Address	Same as above			_
Signature	Kemell V.C.	<u> </u>	Date 1//30/99	_
Full Name of Second Inver	ntor KINZLER	Kenneth		
	Family Name	First Given Name	Second Given Name	
Residence	1403 Halkirk Way, BolAir, M	Maryland 21015		
Citizenship	United States of America			
Post Office				
Address	Same as above			-
Signature			Date	
Full Name of Third Invent	f or	Hans		
TIMO MACIO	Family Name	First Given Name	Second Given Name	
Residence	744 Northcrest Drive, Salt L	ake City, Utah 84103		
Citizenship_	Denmark			
Post Office	_			
Aggress	Same as above			_
Signature Full Name o	s		Date	_
Fourth Inver	itor ANAND	Rakesh		
	Family Name	First Given Name	Second Given Name	
Residence_	62 Grange Way, Sandbach, (Cheshire CW11 9FS England		_
Citizenship_	British			
Post Office	Same as above			
Waress***	Canto as above			_

And we hereby appoint, both jointly and severally, as our attorneys with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected herewith the following attorneys who are all members of the Bar of the District of Columbia, their registration numbers being listed after their names:

Donald W. Banner, Registration No. 17,037; Edward F. McKie, Jr., Registration No. 17,335; William W. Beckett, Registration No. 18,262; Dale H. Hoscheit, Registration No. 19,090; Joseph M. Potenza, Registration No. 28,175; James A. Niegowski, Registration No. 28,331; Joseph M. Skerpon, Registration No. 29,864; Thomas L. Peterson, Registration No. 30,969; Nina L. Medlock, Registration No. 29,673; William J. Fisher, Registration No. 32,133; Thomas H. Jackson, Registration No. 29,808; Patricia E. Hong, Registration No. 34,373; Robert S. Katz, Registration No. 36,402, Brian E. Hanlon, Registration No. 40,449, Sarah A. Kagan, Registration No. 32,141 and Lisa M. Hemmendinger, Registration No. 42,653.

All correspondence and telephone communications should be addressed to: Banner & Witcoff, Ltd., Eleventh Floor, 1001 G Street, N.W., Washington, D.C. 20001-4597, telephone number (202) 508-9100, which is also the address and telephone number of each of the above listed attorneys.

Engine			Date
Al Nume of	YOGELSTEIN	Ben	
The madent—	Family Name	First Given Name	Second Given Name
- oidenne 3	100 Broton Way, Baltimore, Ma	and 21202	
		H TANK X 1200	
tizenskip U ost Office	nitra States of America		
	ovoda se sm		
mature			Date
scond Livento	KNZLER	Kenneth	W .
	Pamily Name	First Given Name	Second Given Name
eridence1	403 Halldric Way, BelAir, Man	Vand 21015	
OR Office	Inited Stores of America		
Zcanbb.	we az apove		
_			
	Se Ville		11-22-0
Ull Nune of	also Of		Date 11-23-9
hird Inventor	ALBERTSEN	Huns	
	Pamily Name	First Given Name	Second Coven Name
ceidence7	44 Northeress Drive, Sait Lake	City Utah 84103 /034	10 VICKEBURG DRIVE
icizenshin L	annark		RTINO, CA 95014
ost Office			121.100 CH 130.17
AddressS	ame aa above		
Signature			v
Pull Name of			Date
יחשרולו לחיבהים	OT ANAND Pame	First Given Name	
			Second Given Name
tesidance	62 Grange Way Sandbach Ch	shire CW11 9FS England	
iozcachip_1	British		
Post Office	Same as above		
2TI-175-04	Same as moste		

LAW DEFICES

BANNEN & WITCOFF, LTD.
1001 a STREET, N.W.
WASHINGTON, D.C. 2000]-4597
(2021 608-8100

And we hereby appoint, both jointly and severally, as our attorneys with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected herewith the following attorneys who are all members of the Bar of the District of Columbia, their registration numbers being listed after their names:

Donald W. Banner, Registration No. 17,037; Edward F. McKie, Jr., Registration No. 17,335; William W: Beckett, Registration No. 18,262; Dale H. Hoscheit, Registration No. 19,090; Joseph M. Potenza, Registration No. 28,175; James A. Niegowski, Registration No. 28,331; Joseph M. Skerpon, Registration No. 29,864; Thomas L. Peterson, Registration No. 30,969; Nina L. Medlock, Registration No. 29,673; William J. Fisher, Registration No. 32,133; Thomas H. Jackson, Registration No. 29,808; Patricia E. Hong, Registration No. 34,373; Robert S. Katz, Registration No. 36,402, Brian E. Hanlon, Registration No. 40,449, Sarah A. Kagan, Registration No. 32,141 and Lisa M. Hemmendinger, Registration No. 42,653.

All correspondence and telephone communications should be addressed to: Banner & Witcoff, Ltd., Eleventh Floor, 1001 G Street, N.W., Washington, D.C. 20001-4597, telephone number (202) 508-9100, which is also the address and telephone number of each of the above listed attorneys.

Signature			Date	
Full Name of First Inventor		Bort		
1 Her III CHOI	Family Name	First Given Name	Second Given Name	
Residence	3700 Breton Way, Baltimore, Ma	ryland 21208		
Citizenship	United States of America			
Post Office				
Address	Same as above			
established Specific				
Signature			Date	
Full Name of				
Second Inven	tor KINZLER Family Name	Kenneth First Given Name	W. Second Given Name	
and the second	Failing Name	Lust quett uans	Second Given Name	
Residence	1403 Halkirk Way, BelAir, Mary	land 21015		
Citizenship	United States of America			
Post Office				
*	Same as above			
Signature			D.e.	
Full Name of	· · · · · · · · · · · · · · · · · · ·		Date	
Third Invento	rALBERTSEN	Hans First Given Name		
	Pamily Name	First Given Name	Second Given Name	
	744 Northcrest Drive, Salt Lake	City, Utah 84103		
Citizenship				
Post Office	Denmark			
	Same as above			
*				
			intlic is	000
Signature Full Name of	rideesh of	rand_	Date 12th November 1	32
Fourth Invent	orANAND	Rekceh	,	
	Family Name	First Given Name	Second Given Name	
Residence	62 Grange Way, Sandbach, Ches	hire CW11 9FS England		
Citizenship_	British			
Post Office Address	Same as above			

LAW OFFICES
BANNER & WITCOFF, LTD.
1001 G STREET, N.W.
WASHINGTON, D.C. 20001-4597
(202) 508-9100

____ <u>B</u>008

Name of	11W4 (1)-(W	son	Date 11 30 6
Loventor	Family Name	Pirot Given Name	Second Olven Name
	174 H. Suraveida Avenue, Salt La	CA City Uph	
nahip <u>U</u> n Office	nited States of America		
	ma sa shaye		
Name of			Date
Dventer_	GRODEN Ramily Name	Pire Given Neges	Second Given Namp
kn∝_67	22 9h Avenue Salt Labo Cler. U	•	
	nited States of Avastics		
Oxfine PullSo	too to spoke		
	•		
Name of			Date
anth Invent	FAMILY Name	Phot Cives Name	Second Green Native
College B	ADIO NA ADIOVE		
mahum			Dute
Name of int Inventor	JOSLYN	Genff	
	panuly Nems	Part Green Name	Seacound Given Navio
	126 To Avenue Sult Lybe Ciry I	lah R41073	
r Olive	Trilled Stoles of America		
drasS	PANE SE SPORE		
mature			Date
il Name of	MARKHAM	Alexander	Food
	Family Name	First Gwen Name	Smooth Green Name
		Chestales England	
izenohip_1 of Office		***	
dresoS	Same as above	***	
			Date
Cardina.			
11 Mamo of	RAKUMURA	Vienthe	
10 anne of		Yunder First Civen Name	Second Goven Name
il Mamo of	RAKUMURA	First Given Name	
onthro di Mano of onth Invento onidence dizenship	Family Namo 1-3-3 Marrivania Kiyose Toky	First Given Name	

The first of the state of the s

The second secon

			Date
Mamo of Laventor	CARLSON	Mary	
M4 4 4 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	CARLSON Family Name	First Olyca Name	Second Given Nume
donne2074.B	Sunnveide Avenue, Sale Labor	in Uch	
mahip United S			
Office			
bers Same	thave		
CAM	m/ mdin	/	Data 11/22/99
Name pl	The same of the sa		
in Investor		Pipes Green Name	Socond Or-un Name
17	Farmy Name	•	
idence 679 9th	Avonue Salt Lake, Sity Ulah	B4102	
issoshin United	Smith of America		
st Clina			
dress Same as	LIMITE		
			►
Il Name of			Dato
Manga of	HEDGE	Philip	
	Patrily Manne	PINE GIVED NAME	Scoops Given Name
widence 7 Rec	berry bise Winsford Cheshing	CW7 3Kg Postend	
incomplie British			
anteri Same i	ie aboye		
			~ .
			La main
ignature			Date
ull Name of	JOSLYN	Grant	
ignature ult Name of ight Inventor	IOSI YN Pamily Namo	Fant Gron Name	Second Given Name
ult Name of ight Inventor	Family Name	Part Gron Name	
ult Name of ight Inventor	Family Name th Avenue, Selt Loke Sity Ute	Part Gron Name	
ult Name of ight Inventor	Family Name	Part Gron Name	
ult Name of ight Inventor 426 7	Family Name th Avenue, Selt Loke Sity Ute	Part Gron Name	
ult Name of ight Inventor 426 7	Family Name The Avenue, Selt Lake City Ute Scales of America	Part Gron Name	
ill Nume of ight Inventor Caldware 426 T Caldware 426 T Caldware 426 T Caldware 426 T Caldware 426 T Caldware 5 Same	Family Name The Avenue, Selt Lake City Ute Scales of America	Part Gron Name	Second Given Name
ill Name of ight Inventor 426 7	Family Name The Avenue, Sell Lake City Und Scales of America As shows	Part Gron Name	Second Given Name
ill Nume of ight Inventor Caldware 426 T Caldware 426 T Caldware 426 T Caldware 426 T Caldware 426 T Caldware 5 Same	Family Name The Avenue, Sell Lake City Ute Scales of America As shown MARKHAM	Pint Gron Name	Second Grean Name Date Prod
ill Name of special Name of Same	Family Name The Avenue, Sell Lake City Und Scales of America As shows	Pint Gron Name	Second Given Name
ill Name of ight Inventor 426 7 coldman 426 7 col	Family Name The Avenue, Sell Lake City Und Scales of America MARKHAM Family Name	Pint Gron Name h 14191 Alreander First Grozn Name	Second Grean Name Data
ill Name of ight Inventor 426 7 ight Inventor 426 7 ight Difference 5 5 mm defense 5 mm fruit Name of Numb Inventor 6 Residonec 25	Family Name The Avenue, Sell Lake City Under Solver of America As shown MARKHAM Family Name Though Red Lake Sponsory, Care	Pint Gron Name	Second Grean Name Data
ill Name of ight Inventor 426 7 initiation 426	Family Name The Avenue, Sell Lake City Under Solver of America As shown MARKHAM Family Name Though Red Lake Sponsory, Care	Pint Gron Name h 14191 Alreander First Grozn Name	Second Grean Name Date Prod
ill Name of ight Inventor 426 7 ight Inventor 426 7 ight Difference 5 5 mm defense 5 mm fruit Name of Numb Inventor 6 Residonec 25	Family Name The Avenue, Sell Lake City Under States of America A show MARKHAM Family Name Rooth Red Lane Scenarry, Catalah	Pint Gron Name h 14191 Alreander First Grozn Name	Second Grean Name Data
ull Name of ight Inventor ight Inventor ight Inventor ight Inventor ight Name of Name ight Name of Name of Name Residence 25 Citizenship Reidi Post Office	Family Name The Avenue, Sell Lake City Under States of America A show MARKHAM Family Name Rooth Red Lane Scenarry, Catalah	Pint Gron Name h 14191 Alreander First Grozn Name	Second Grean Name Data
ull Name of ight Inventor coldman 426 7 coldman 5 coldman 5 coldman 6 coldman 7 coldman	Family Name The Avenue, Sell Lake City Under States of America A show MARKHAM Family Name Rooth Red Lane Scenarry, Catalah	Pint Gron Name h 14191 Alreander First Grozn Name	Deta
ull Name of ight Inventor coldman 426 7 coldman 5 coldman 5 coldman 6 coldman 7 coldman	Family Name The Avenue, Sell Lake City. Und Scales of America As above MARKHAM Family Name Crooth Red Lace Sconney, Contact and above	Pint Gron Name h 14191 Alreander First Grozn Name	Second Grean Name Data
ill Name of ight Inventor ight Inve	Family Name The Avenue, Sell Lake City. Und Scales of America As above MARKHAM Family Name Crooth Red Lace Sconney, Contact and above	Pient Gron Name Altomater Flirst Grown Name Sec. Chashire, Hagland	Date
ull Name of ight Inventor coldman 426 7 coldman 5 coldman 5 coldman 6 coldman 7 coldman	Family Name The Avenue, Sell Lake City Under States of America A show MARKHAM Family Name Rooth Red Lane Scenarry, Catalah	Alcounder First Given Name Alcounder First Given Name Stee Chashire, Bugland	Second Given Name Date Prod Second Given Name
ill Name of ight Inventor Ight Inve	Family Name The Avenue, Sell Lake City Under States of America MARKHAM Family Name Booth Bed Lane Generary, Carlo And Andrew Partilly Name	Alcounder First Green Name Chashire, Bugland Yuguka Figut Green Name	Date
ill Name of ight Inventor ight Inventor ight Inventor ight Differ Same Superior	Family Name She avenue, Salt Lake City. Und Scales of America MARKHAM Formity Name Choole Bod Lane Scansory. Con She above Naturally Name Carrier Name Salt Margyana, Kivose Tokyo	Alcounder First Green Name Chashire, Bugland Yuguka Figut Green Name	Date Date Date Date
ill Name of ight Inventor Ight Inve	Family Name She avenue, Salt Lake City. Und Scales of America MARKHAM Formity Name Choole Bod Lane Scansory. Con She above Naturally Name Carrier Name Salt Margyana, Kivose Tokyo	Alcounder First Green Name Chashire, Bugland Yuguka Figut Green Name	Date Date Date Date

Law offices
Banner & Witcom, Ltd.
1007 & Etreet, N.W.
WASHINGTON, D.C. 10007-4587
12021 E02-8100

Signature	•		Date
Signature Pull Name of			
Fifth Inventor	CARLSON Family Name	Mary First Given Name	Second Given Name
	•		•
Residence	2074 B. Sunnyside Avenue, Salt L	ake City. Utah	
Citizenship Post Office	United States of America		
	Same as above		
•			
Signature			Date
Full Name of	rGRODEN	Јоаппа	
SIAUI IIIVOIKO	Family Name	First Given Name	Second Given Name
Residence	629 9th Avenue, Salt Lake City, U	Jtah 84103	
Post Office			
Address	Same as above		
	1110	1/ 1	
Signature	This !	NOT	Date 12H November 199
Full Name of		Photics	
Seventh Inve	Family Name	Philip First Given Name	Second Given Name
n • •	•		·
		tire CW7 3EA England	
Citizenship Post Office	British		
	Same as above		
Limited Lim			
inner;			
Signature			Date
Full Name of		Con #	
Eight Invento	or JOSLYN Family Name	Geoff First Given Name	Second Given Name
Andrews Constitution	•		
Residence	426 7th Avenue, Salt Lake City, I	Jtah 84103	
Citizenship_	United States of America		
Post Uilice	Same as above		
71001055	Same as accept		
To go			
			. .
Signaturc_ Puli⊧Name o:	f	· · · · · · · · · · · · · · · · · · ·	Date
Ninth Invent	or MARKHAM	Alexander	Fred
	Family Name	First Given Name	Second Given Name
Residence	25. Booth Bed Lane, Goostrey, C.	rewe. Cheshire. England	
Citizenship_	British		
Post Office Address	C		
Wooless	Same as above		
Signature Full Name of			Date
Tenth Invent	orNAKUMURA	Yusuka	
	Family Name	First Given Name	Second Given Name
Residence	1-43-3 Matsuyama, Kivose Tokyo	204 Japan	
Citizenship_			
Post Office			
Address	Same as above		

CARLSON FAMILY NAME THE SAMPLING ANDRES THE SAMPL	Miny Prot Group Refer	Second Given Water
Family Manne 114 B. Sennyaids Armore Sale Loke C nized States of Ampeica mana shorts	First Go-un Nation	Second Corons (Value
114 B Separation Arrange Sale Lake Control States of Amprica	First Go-un Nation	
nied States of America		
nied States of America		
Dear shorts		
DEN POYE		
GRODPN Ramilly M.		
GRODPN Ramilly M		
GRODPN Parally M		
GRODEN Formula M		Dubr
Peners Dat Toler	Janaha	
V	Pare Giron Masse	
9 9th Avenue Sult Labe City Ush B		Second Green Name
no ne shows		
•		
		<u>.</u>
r REDGR	Need was	Dec
	Tollo.	
•	That online Milit	Second Gaven Name
Monkey Risc Winsfard Chashing	17 3PA Papland	
rosh		
DE 86 Above		
General Dala		,
		- 11/27
905LYN		1 Page 1 - 3
Pamily Name	First Govern Warner	
& Th Assessed Sult Laboration Visit to		Secreta Carrent Name
	ENCINIT.	AS CA 92024
and are above		17027
		- -
		Date:
MARKHAM	Almondor	Food .
	First Geron Name	Strong Lives Name
Booth Red Lane Godgey Come !	Chestine Parland	
or as above		
		
		_
MARTIMITO		Dans
Patrilly Narran	Pitel Corn Name	
·		Second Given Name
1 		
3-3 Marsuvana Kivaso Tokyo 204 J.	ion.	فلالمها و مساحد
	Bookery Rise: Wineferd Chashire Critish Bookery Rise: Wineferd Chashire Critish Family Name In The Assembly Salk Lake City Utah Mandard Stakes of America Mank Ham Family Name Books Red Lane Godgesy Cresses HAKUMURA	REDGE Partity Name Provided Schools Charles CW7 3PA Particol Respective Rise. Whatfand. Charles CW7 3PA Particol Respective Rise. Whatfand. Charles CW7 3PA Particol Respective Rise. Whatfand. Charles Respective Respec

Law offices
Banned & Wittoner, Ltd.
1007 G Stireet, N.W.
Washineton, D.C., 2000 4507
12071 540.0100

			Date
Signature Full Name of			To the same of the
Fifth Inventor	CARLSON Family Name	Mary First Given Name	Second Given Name
	•		
Residence 2074	LE Sunnyside Avenue, Salt Lal	se City Utah	
Citizenship Unit	ed States of America		
Post Office			
AdoressSame	s as above		
Signature Full Name of		Chinario, and Chinario and Chin	Date
Full Name of Sixth Inventor	GRODEN	loanes	
	Family Name	First Given Name	Second Given Name
Residence 629	9th Avenue, Salt Lake City, Ut	th 84103	
Conzonship <u>Unit</u> Post Office	ed Suites Of Affiches	All the second s	
Address Sam	c as above		
Signature			Date
Full Name of			47 04 F
Seventh Inventor	HEDGE Family Name	Philip Pirst Given Name	Second Given Name
	•		
Residence7_R	lookery Rise, Winsford, Cheshi	e CW7 3BA England	
Citizenship_Briti	sh		
Post Office Address Sam	e as above		
Signature Full Name of			Date
Full Name of Eight Inventor	JOŠLYN	Geoff	
	JOSLYN Family Name	First Given Name	Second Given Name
Residence 426	7th Avenue, Salt Lake City, Ut	ah 84103	
Post Office	ted States of America		
AddressSam	e as above		
	N /		2
	May May W		Date November 199
Signature Full Name of	111 000.		Date! NOVELLEGE POUR
Ninth Inventor	MARKHAM	Alexander	Fred
	Family Name	First Given Name	Second Given Name
Residence 25.	Booth Bed Lanc, Goostrey, Cre	we, Cheshire, England	
Citizonship_Briti			
Post Office			
Address Sam	e as above		and the same of th
Signature			Date
Full Name of	** * *** * * * * * * * * * * * * * * *		Datc
Tenth Inventor	NAKUMURA Family Name	Yusuka First Given Name	Second Given Name
Danislana	•		amotic Atten Hanne
Residence 1-4	3-3 Matsuyama, Kiyose Tokyo 2	ALP Japan	
Citizenship Japa	nesc		
Post Office Address Sam	e as above		

mantee			Datc
manire Il Name of			
th Inventor	CARLSON Family Name	Mary First Given Name	Second Given Name
		City, Utah	
	States of America		
t Office			
dress <u>Same</u> s	as above		
mature			Date
ll Name of		*	
th Inventor	GRODEN Family Name	Joanna First Given Name	Second Given Name
	- •		
izenship United	States of America		
st Office Idress Same	as above		
			Date
II Nome of			Datt
venth Inventor	HEDGE Family Name	Philip	0 VC' V
	Family Name	First Given Name	Second Given Name
sidence 7. Ro	okery Rise Winsford Cheshire	CW7 3EA England	
tizenship <u>Britis]</u>			
	as above		
Rendered			
To the state of th			_
gnature			Date
ill Name of ght Inventor	JOSLYN	Geoff	
	Family Name	First Given Name	Second Given Name
	7th Arranga Solt I also City I litch	84103	
itizonship <u>Unite</u> ost Office	d States of America		
ddress Same	as abovo		
and the second			
T. C.			
gnature			Date
il Name of		-	- ·
inth Inventor	Family Name	Alexander First Given Name	Fred Second Given Name
	•		
esidence 25 E	Booth Bed Lane, Goostrey Crew	e, Cheshire, England	
itizenship <u>Britis</u>	ih		
ost Office .ddress <u>Same</u>	as above		
		_	
. /	121. A. C		11/13/99
ignature ull Name of	Jun (Mel		Date / 10 / /
enth Inventor	NAKUMURA	Yusuks	
	Family Name	First Given Name	Second Given Name
esidence <u>1-43</u>	-3 Matsuyama, Kiyose Tokyo 20	4 Јарап	
itizonship <u>lapar</u>	nese		
ost Office	an about		

11/23/99	TUE 14:03 FAX 601 881 7838	Unturtech transfer		2007
Signature Poll Name of	NW	- 1/	Dem 11/24/99	
Elmousy Inca Lot Mares or	Party Nema	Pire Givait Namo	Second Given Name	
Cidacratiq	3704 S. 2140 S. Belt Labo Chy Lingh Mild Visited Sector of America Strong of America			The same of the sa
Pall Name of	WHITE Farely Name	Navana Funt Closy Name	Dates	_
	711 12th Avenue, Rell Labe Cler, Unit PAI United States of America	• • • • • • • • • • • • • • • • • • • •		

Law offices
Benner & Wittoff, Ltd.
1001 & Stelly, N.W.
Warnington, R.C. 20001-4887
[201] Bod-ni &

The Target of Ta

11/24/99	WED	13:0	7 FAX	801	581	753	8
11/17	/99,	WED	14:07	FAX	86	81	7538

Uofu/TECH TRANSFER

Uofu/Tech Transfer

□ 011 r. □ 1 - _ _ 2007

		Detr
Signature Signature Signature Figure 101 Figure 10	First Given Name	Second Given Name
Besidence 1794 S 2140 H Salt Lake City, Unit 84100		
Citizenship United States of America Post Office Address Same as above		- 11
Ray White		Date 11 19 19
Signature Full Name of Techh Investor Family Name Family Name	Plate Civon Namo	Second Given Name
Muddenco 711 18th Avenue Sale Lake City Utah 841	»	
Cingenship United States of America		
Post Office		

SEQUENCE LISTING

```
( 1 ) GENERAL INFORMATION:

( i i i ) NUMBER OF SEQUENCES: 102

( 2 ) INFORMATION FOR SEQ ID NO:1:

( i ) SEQUENCE CHARACTERISTICS:
( A ) LENGTH: 9606 base pairs
( B ) TYPE: nucleic acid
( C ) STRANDEDNESS: double
( D ) TOPOLOGY: linear

( i i ) MOLECULE TYPE: cDNA

( v i ) ORIGINAL SOURCE:
( A ) ORGANISM: Homo sapiens

( v i i ) IMMEDIATE SOURCE:
( B ) CLONE: DP2.5(APC)

( i x ) FEATURE:
( A ) NAME/KEY: CDS
```

-continued

1	R	LOC	ATION:	348562

DJIHELDO LILOS

		(B) LOCA	TION: 34.	.8562											
	(xi)	SEQUEN	ICE DES	CRIPTION	: SEQ ID	NO:1:										
GGAC	TCGC	SAA A	TGAC	GTCC	A AG	GGTA	GCCA	AGG						Тут	GAT Asp	5 4
CAG Gln	TTG Leu	TTA Leu 10	AAG Lys	CAA Gln	GTT Val	GAG Glu	GCA Ala 15	CTG Leu	AAG Lys	ATG Met	GAG Glu	AAC Asn 20	T C A S e r	AAT Asn	CTT Leu	102
				GAA Glu												150
				ATG Met												198
				ATG Met 60												2 4 6
Leu	Lys	Glu	Leu 75	AAC Asn	Leu	Asp	Ser	8 c r	Asn	Phe	Pro	Gly	Val 85	Lys	Leu	294
Агд	Ser	Lys 90	Met	TCC Ser	Leu	Агд	S e r 9 5	Туг	Gly	Ser	Arg	G 1 u 100	Gly	Ser	V a l	3 4 2
Ser	Ser 105	Arg	Ser	GGA Gly	G 1 u	Сув 110	Ser	Pro	V a 1	Pro	Met 115	Gly	Ser	Phe	Рго	390
Arg 120	Arg	Gly	Phe	GTA Val	Asn 125	G 1 y	Ser	Arg	Glu	Ser 130	Thr	G 1 y	Туг	Leu	Glu 135	438
Glu	Leu	Glu	Lys	GAG Glu 140	Arg	Ser	Leu	Leu	Leu 145	Ala	Asp	Leu	Asp	Lys 150	Glu	486
Glu	Lys	G 1 u	Lys 155		Trp	Tyr	Туг	A 1 a 160	G1n	Leu	Gln	Asn	Leu 165	Thr	Lys	534
Агд	Ile	Asp 170	Ser	CTT Leu	Рго	Leu	Thr 175	Glu	Asn	Phe	Ser	Leu 180	Gln	Thr	Asp	5 8 2
Leu	Thr 185	Arg	Arg	CAA Gln	Leu	G 1 u 1 9 0	Туг	Glu	Ala	Arg	G 1 n 195	Ile	Агд	Val	Ala	630
	Glu			Leu												678
				AGA Arg 220											Ile	7 2 6
				CAG Gln										Ser		774
			G 1 u	ACC Thr				Asp								8 2 2
		va 1		GAA Glu												870
	Th			A ATG Met							V a 1					918

								-coi	ntinued	1						
								CTG Leu								966
								T C A S e r 3 2 0								1 0 1 4
GAT Asp	GAT Asp	ATG Met 330	T C G S e r	C G A A r g	ACT	TTG Leu	CTA Leu 335	GCT Ala	ATG Met	TCT Ser	AGC Ser	T C C S e r 3 4 0	CAA Gln	GAC Asp	AGC Ser	1062
								TGT Cys								1110
TTA Leu 360	CAT His	GGC Gly	AAT Asn	GAC Asp	A A A L y s 3 6 5	GAC Asp	TCT	GTA Val	TTG Leu	TTG Leu 370	GGA Gly	AAT Asn	TCC	CGG Arg	GGC G1 y 3 7 5	1 1 5 8
								AGT Ser								1206
								GGC G1y 400								1 2 5 4
								TAC								1 3 0 2
								G A C A s p								1350
								Pro								1398
Leu	Ser	Phe	Азр	G1 u 460	Glu	His	Arg	CAT His	A 1 a 4 6 5	Met	Asn	Glu	Leu	G 1 y 4 7 0	G 1 y	1446
Leu	Gln	Ala	I 1 e 4 7 5	Ala	Glu	Leu	Leu	CAA Gln 480	Val	Asp	Сув	Glu	Met 485	Туr	Gly	1494
Leu	Thr	Asn 490	Asp	His	Туг	Ser	I 1 e 4 9 5		Leu	Агд	Агд	T y r 5 0 0	Ala	Gly	Met	1542
Ala	Leu 505	Thr	Asn	Leu	Thr	Phe 510	Gly	GAT Asp	Val	Ala	A s n 5 1 5	Lys	Ala	Thr	Leu	1590
C y s 5 2 0	Ser	Met	Lys	G 1 y	C y s 5 2 5	Met	Arg	GCA Ala	Leu	V a 1 5 3 0	Ala	Gin	Leu	Lys	S e r 5 3 5	1638
Glu	Ser	Glu	Asp	Leu 540	Gln	Gln	V a 1	Ile	A 1 a 5 4 5	Ser	V a 1	Leu	Агд	A s n 5 5 0		1686
Ser	Trp	Arg	A 1 a 5 5 5	Asp	Vai	Asn	Ser	L y s 5 6 0	Lys	Thr	Leu	Arg	G 1 u 5 6 5	V a 1		1734
Ser	Val	L y s 5 7 0	Ala	Leu	Met	Glu	Cys 575	Ala	Leu	Glu	Va1	L y s 5 8 0	Lys	Glu		1782
Thr	Leu 585	Lys	Ser	Val	Leu	Ser 590	Ala	Leu	Trp	Asn	Leu 595	Ser	Ala	His		1830
	Glu					Ile					Gly				TTT Phe 615	1878

				31										<i>5</i> 0		
								-coi	ntinued	Į						
												4.0.5		200	4 T T	1006
				CTT Leu												1926
Leu	Val	Gly	lbr	620	1 11 1	1 9 1	Arg	3 C I	625	1 11 1	ASI	1 11 1	Leu	630	110	
				020												
				GGT												1974
I 1 e	Glu	Ser		Gly	Gly	I 1 c	Leu		Asn	Val	Ser	Ser		Ile	Ala	
			635					640					6 4 5			
4.0.4	A A T	GAG	GAC	CAC	A G G	$C \Delta \Delta$	A T C	СТА	AGA	GAG	AAC	AAC	тст	CTA	CAA	2022
				His												
		650			0		6 5 5		·			660				
				CAC												2070
Thr		Leu	Gln	His	Leu	Lys 670	Set	His	Ser	Leu	675	116	Val	Ser	Asn	
	665					0 7 0					0 7 3					
GCA	TGT	GGA	ACT	TTG	TGG	AAT	CTC	TCA	GCA	AGA	AAT	CCT	AAA	GAC	CAG	2 1 1 8
				Leu											Gln	
680					685					690					695	
			m a a		4 75 0			C T T	400	4 T.C	CTC	4 4 6		СТС	A T T	2166
GAA	GCA	TIA	Ten	GAC Asp	Mat	GUG	Ala	Val	Set	Met	Len	T. v e	ARC	Len	I 1 e	2100
014	A. a	Ltu		700	1410 1	0.,			705			-,-		710		
				AAA												2214
His	Ser	Lys		Lys	Met	110	Ala		Gly	Ser	Ala	Ala		Leu	Arg	
			715					720					7 2 5			
A A T	CTC	ATG	GCA	AAT	A G G	ССТ	aca	AAG	TAC	AAG	GAT	acc	AAT	ATT	ATG	2 2 6 2
				Asn												
		730					735	•	•	•	-	740				
				AGC												2310
Ser		Gly	Ser	Ser	Leu	750	Ser	Leu	H 1 8	Vai	A 1 g	Lys	GIn	Lys	AIA	
	7 4 5					130					733					
CTA	GAA	GCA	GAA	TTA	GAT	GCT	CAG	CAC	TTA	TCA	GAA	ACT	TTT	GAC	AAT	2 3 5 8
Leu	Glu	Ala	Glu	Leu	Asp	Ala	Gln	His	Leu	Ser	G 1 u	Thr	Phe	Asp	Asn	
760					765					770					775	
						~	~ ~ .	m a m	~	ОСТ					C 4 C	2406
				AGT Ser												2400
116	Asp	ASH	Dea	780	110	L, s	AIA	501	785		•••	٠,,,	0	790		
				TAT												2 4 5 4
Lys	Gln	Ser		Туг	Gly	A s p	Туг		Phe	Asp	Thr	Asn		His	Asp	
			795					800					805			
GAT	AAT	AGG	TCA	GAC	AAT	TTT	AAT	ACT	GGC	AAC	ATG	ACT	GTC	CTT	TCA	2502
Asp	Asn	Arg	Ser	Asp	Asn	Phe	Asn	Thr	G 1 y	Asn	Met	Thr	Va1	Leu	Ser	
-		810					8 1 5					820				
				ACT												2550
Pro	1 y r 8 2 5		Asn	Thr	111	8 3 0		PIO	SEI	Ser	835		361	Arg	Gry	
	023					030					0,00					
AGC	TTA	GAT	AGT	TCT	CGT	TCT	GAA	AAA	GAT	AGA	AGT	TTG	GAG	AGA	GAA	2598
Ser	Leu	Asp	Ser	Ser	_		Glu	Lys	Asp			Leu	G 1 u	Arg		
8 4 0					8 4 5					850					8 5 5	
000				CTA			T 4 C	C 4 T	C C A	GC 4	4.0.4	G A A	A A T		GGA	2646
				Leu												2040
,	,		,	860			- , -		865					870		
				CGA												2694
Tht	Ser	Ser		Arg	Gly	Leu	Gla			Thr	Thr	Ala			116	
			875	•				880					8 8 5	•		
acc	. AAA	GTC	ATO	GAA	GAA	GTG	TCA	GCC	ATT	CAT	ACC	тст	CAC	GAA	GAC	2742
				Glu												
	•	890					895					900				
_																***
				TCT Ser											AGA	2790
АГЕ	905		GI	, ser	IDI	910		гел	пів	C y s	915		us]	, ,,,,,	. vi A	
	, , ,	•				0										
				A AGA												2838
		a Leu	Arı	g Arg			Ala	Ala	His			Ser	A 8 1	Thr		
920)				9 2 5	5				930)				935	

				39										70		
								-cont	inued							
AAT 3				=	a	445 7	5.6.4	A A T	. aa	4 C 4	тат	тет	A T.G	ССТ	тат	2886
AAT I	TTC	ACT	AAG	Ser	GIR	AAI I	l CA Se r	Asn A	Are	Thr	Cvs	Ser	Met	Pro	Tyr	2000
ASI I	пс	1 11 1	Lys	940	010				9 4 5		- , -			950	-	
										a					OT C	2934
GCC A	AAA	TTA	GAA	TAC	AAG	AGA T	CT	TCA A	AAT	GA1	SCI	Len	AAI	Ser	Val	2934
Aial	Lys	Leu	955	1 9 1	Lys	AI g	, с 1	960		кор	501	200	965			
AGT A	AGT	AAT	GAT	GGT	TAT	GGT A	AAA	AGA (GGT	CAA	ATG	AAA	CCC	TCG	ATT	2982
Ser S	Ser	Asn 970	Asp	Gly	Tyr	Gly I	. у s 975	Arg (i I y	GIn	Met	980	PIO	361	116	
		9/0				•	,,,					, , ,				
GAA :	тсс	TAT	TCT	GAA	GAT	GAT (G A A	AGT .	AAG	TTT	TGC	AGT	TAT	GGT	CAA	3030
		Туг	Ser	G 1 u	Аsр	Asp (31 u	Ser 1	Lys	Phe	Cys	Ser	Туг	Gly	Gln	
!	985					990					995					
TAC	CCA	GCC	GAC	CTA	GCC	CAT	AAA	ATA	CAT	AGT	GCA	AAT	CAT	ATG	GAT	3078
Tyr	Pro	Ala	Asp	Leu	Ala	His	Lys	Ile !	His	Ser	Ala	Asn	H i s	Met	Asp	
1000					1005	5				1010)				1015	
GAT	A A T	GAT	GGA	GAA	СТА	GAT .	A C A	CCA	АТА	ААТ	TAT	AGT	СТТ	AAA	TAT	3126
Asp.	Asn	Asp	Gly	Glu	Leu	Asp	Thr	Pro	ile	Asn	Туг	Ser	Leu	Lys	Туг	
		-	·	102					1025	;				103	0	
			~ . ~			TCT	~~ .		~	AGT	CCT	TCA	CAG	A A T	GAA	3174
TCA	GAT	GAG	Gin	IIG	AAC	Ser	GUA	ATO	Gin	Ser	Pro	Ser	Gln	Asn	Glu	5.7.4
361	дър	0.0	103				· . ,	1040					104	5		
AGA	TGG	GCA	AGA	ccc	AAA	CAC	ATA	ATA	GAA	GAT	GAA	ATA	AAA	CAA	AGT	3 2 2 2
Arg	Trp	Ala 105		Pro	Lys	His	1055		GIU	Asp	GIU	106	О	O I I	361	
GAG	CAA	AGA	CAA	TCA	AGG	AAT	CAA	AGT	ACA	ACT	TAT	CCT	GTT	TAT	ACT	3 2 7 0
G 1 u			Gln	Ser	Arg	Asn	Gln	Ser	Thr	Thr	Tyr 107:	Pro	Val	Туг	Thr	
	106	5				1070					107.	3				
GAG	AGC	ACT	GAT	GAT	AAA	CAC	CTC	AAG	TTC	CAA	CCA	CAT	TTT	GGA	CAG	3 3 1 8
Glu	Ser	Thr	Asp	Asp	Lys	His	Leu	Lys	Phe	G 1 n	Рго	His	Phe	Gly	Gln	
1080					108	5				109	0				1095	
CAG	GAA	тст	GTI	тст	CCA	TAC	AGG	TCA	CGG	GGA	GCC	AAT	GGT	TCA	GAA	3 3 6 6
Gln	Glu	Cys	Val	Ser	Pro	Tyr	Агд	Ser	Arg	Gly	Ala	Asn	G 1 y	Ser	Glu	
				110					110	5				1 1 1	0	
			ОТ		тст	AAT	CAT	GGA	ATT	A A T	CAA	AAT	GTA	AGG	CAG	3 4 1 4
The	Asn	Arg	Val	Glv	Ser	Asn	His	Gly	I 1 e	Asn	Gln	Asn	Val	Ser	Gln	
			111					1120					1 1 2	5		
						~ . ~	T A T	~	GAT	CAT	4 4 6	CCT	A.C.C		TAT	3 4 6 2
TCT	TTG	Cva	GI	A GAA	GA1	GAC Asp	Tvr	Gln	Asp	Asp	Lvs	Pro	Thr	Asn	Tyr	3402
361	Ltu	113			11 0 p	,	113				- , -	114	0		•	
																2510
AGT	GAA	CGI	TAC	CTCT	GAA	GAA Glu	GAA	CAG	CAT	GAA	GAA	GAA	GAG	AGA	Pro	3510
Ser	114		; 1 y :	r ser	Giu	1150		GII	1113	010	115	5	0.0		,	
ACA	AAT	LAT	AG	C ATA	AAA	TAT	AAT	GAA	GAG	AAA	CGT	CAT	GTC	GAT	DAO 1	3 5 5 8
		Туг	S c	rile	Lys 116	Туг	Asn	Glu	Glu	Lys 117	Arg	n 1 8	val	ASĮ	1175	
1160											-					
CCT	ATI	GA1	AT 1	T AGI	TTA	AAA	TAT	GCC	ACA	GAT	ATT	CCT	TCA	TCA	A CAG	3606
Pro	Ile	Asp	Ту	r Ser	Leu	Lys	Туг	Ala	Thr	Asp	I 1 e	Pro	Ser	Se 1	Gln	
				1 1 8	U				118	J				11;	, •	
AAA	CAC	TC	A TT	T TCA	TTC	TCA	AAG	AGT	TCA	тст	GGA	CAA	AG	AG:	T AAA	3 6 5 4
Lys	Gla	Se	r Ph	e Ser	Phe	Ser	Lуs	Ser	Ser	Ser	Gly	Gln	Sei	Se	r Lys	
			1 1	9 5				120	0				12) 5		
ACC	G A A	. C .	т ат	с тег	тса	AGC	AGT	GAG	AAT	ACG	TCC	ACA	cc	г тс.	A TCT	3702
Thr	Gli	Hi	s Me	t Sei	Ser	Ser	Ser	Glu	Asn	Thr	Ser	Thr	Pro	Se	r Ser	
		1 2					121					1 2 2	0			
			a	a c	2 4 4 7		CTC		CC 4		י דרי	60	CA	G AG	T AGA	3750
AAT Asn	A 1	a Lv	or A.G. s A.r	g Gl	AAI Ast	Gln	Leu	His	Pro	Ser	Ser	Ala	G1:	n Se	r Arg	2,20
	1 2 3			J	••	123					123	5			•	
	_	_	_						m ~ -					T • T	T 440	3798
AGT	GG:	r CA	G CC	T CAA	A AA(GCT Ala	G C C	ACT The	Lec Lec	AAA	a Ull leV i	Set	Se	r II	T AAC	3170
124		, 01	m t.t	J J 1	124				- , ,	125					1255	
	-					-				_						

		continued	
		-continued	
			AT ACT CCA ATA TGT TTT 3846 sp Thr Pro Ile Cys Phe 1270
			CA GCT GAA GAT GAA ATA 3894 er Ala Glu Asp Glu Ile 1285
Gly Cys A	AT CAG ACG ACA sn Gln Thr Thr 290	CAG GAA GCA GAT T Gln Glu Ala Asp S 1295	CT GCT AAT ACC CTG CAA 3942 er Ala Asn Thr Leu Gln 1300
ATA GCA G Ile Ala G 1305	AAA ATA AAA GGA lu Ile Lys Gly	AAG ATT GGA ACT A Lys lle Gly Thr A 1310	GG TCA GCT GAA GAT CCT 3990 rg Ser Ala Glu Asp Pro 1315
		Val Ser Gla His P	CT AGA ACC AAA TCC AGC 4038 ro Arg Thr Lys Ser Ser 330 1335
			CA GCC AGG CAC AAA GCT 4086 er Ala Arg His Lys Ala 1350
			CC AAA AGT GGT GCT CAG 4134 er Lys Ser Gly Ala Gln 1365
Thr Pro I			AG GAG ACC CCA CTC ATG 4182 1 n Glu Thr Pro Leu Met 1380
			AT AGT TTT GAG AGT CGT 4230 sp Ser Phe Glu Ser Arg 1395
TCG ATT C Ser Ile A 1400	GCC AGC TCC GTT Ala Ser Ser Val 140	Gln Ser Glu Pro C	GC AGT GGA ATG GTA AGT 4278 ys Ser Gly Met Val Ser 410 1415
			GC CCT GGA CAA ACC ATG 4326 er Pro Gly Gln Thr Met 1430
			CT CCT CAA ACA GCT CAA 4374 ro Pro Gla Thr Ala Gla 1445
The Lys.			CT ACT GCT GAA AAG AGA 4422 ro Thr Ala Glu Lys Arg 1460
	Gly Pro Lys Gln		GCT GCA GTT CAG AGG GTC 4476 la Ala Val Gin Arg Val 1475
		Asp Thr Leu Leu F	CAT TTT GCC ACA GAA AGT 451 lis Phe Ala Thr Glu Ser 1490 1495
ACT CCA Thr Pro	GAT GGA TTT TCT Asp Gly Phe Ser 1500	TGT TCA TCC AGC C Cys Ser Ser Ser I 1505	erg AGT GCT CTG AGC CTC 456 en Ser Ala Leu Ser Leu 1510
			TTA AGA ATA ATG CCT CCA 461 Leu Arg Ile Met Pro Pro 1525
Val Gin	GAA AAT GAC AAT Glu Asn Asp Ass 1530	GGG AAT GAA ACA G Gly Asa Glu Thr G 1535	GAA TCA GAG CAG CCT AAA 466 Glu Ser Glu Gln Pro Lys 1540
	Asn Glu Asn Gla		GAA AAA ACT ATT GAT TCT 471 Glu Lys Thr Ile Asp Ser 1555
		Asp Ser Asp Asp	GAT GAT ATT GAA ATA CTA 475 Asp Asp Ile Glu Ile Leu 1570 1575

-continued	
GAA GAA TGT ATT ATT TCT GCC ATG CCA ACA AAG TCA TCA CGT AAA GGC Glu Glu Cys Ile Ile Ser Ala Met Pro Thr Lys Ser Ser Arg Lys Gly 1580 1585	4806
AAA AAG CCA GCC CAG ACT GCT TCA AAA TTA CCT CCA CCT GTG GCA AGG Lys Lys Pro Ala Gln Thr Ala Ser Lys Leu Pro Pro Pro Val Ala Arg 1595 1600 1605	4 8 5 4
AAA CCA AGT CAG CTG CCT GTG TAC AAA CTT CTA CCA TCA CAA AAC AGG Lys Pro Ser Gln Leu Pro Val Tyr Lys Leu Leu Pro Ser Gln Asn Arg 1610 1615 1620	4902
TTG CAA CCC CAA AAG CAT GTT AGT TTT ACA CCG GGG GAT GAT ATG CCA Leu Gln Pro Gln Lys His Val Ser Phe Thr Pro Gly Asp Asp Met Pro 1625 1630 1635	4950
CGG GTG TAT TGT GTT GAA GGG ACA CCT ATA AAC TTT TCC ACA GCT ACA Arg Val Tyr Cys Val Glu Gly Thr Pro Ile Asn Phe Ser Thr Ala Thr 1640 1645 1650	4998
TCT CTA AGT GAT CTA ACA ATC GAA TCC CCT CCA AAT GAG TTA GCT GCT Ser Leu Ser Asp Leu Thr Ile Glu Ser Pro Pro Asn Glu Leu Ala Ala 1660 1670	5046
GGA GAA GGA GTT AGA GGA GGA GCA CAG TCA GGT GAA TTT GAA AAA CGA Gly Glu Gly Val Arg Gly Gly Ala Gln Ser Gly Glu Phe Glu Lys Arg 1675 1680 1685	5094
GAT ACC ATT CCT ACA GAA GGC AGA AGT ACA GAT GAG GCT CAA GGA GGA Asp Thr Ile Pro Thr Glu Gly Arg Ser Thr Asp Glu Ala Gln Gly Gly 1690 1695 1700	5 1 4 2
AAA ACC TCA TCT GTA ACC ATA CCT GAA TTG GAT GAC AAT AAA GCA GAG Lys Thr Ser Ser Val Thr Ile Pro Glu Leu Asp Asp Asn Lys Ala Glu 1705 1710	5190
GAA GGT GAT ATT CTT GCA GAA TGC ATT AAT TCT GCT ATG CCC AAA GGG Glu Gly Asp Ile Leu Ala Glu Cys Ile Asn Ser Ala Met Pro Lys Gly 1720 1735	5 2 3 8
AAA AGT CAC AAG CCT TTC CGT GTG AAA AAG ATA ATG GAC CAG GTC CAG Lys Ser His Lys Pro Phe Arg Val Lys Lys Ile Met Asp Gln Val Gln 1740 1745 1750	5286
CAA GCA TCT GCG TCG TCT TCT GCA CCC AAC AAA AAT CAG TTA GAT GGT Gln Ala Ser Ala Ser Ser Ser Ala Pro Asn Lys Asn Gln Leu 'Asp Gly 1755 1760 1765	5 3 3 4
AAG AAA AAG AAA CCA ACT TCA CCA GTA AAA CCT ATA CCA CAA AAT ACT Lys Lys Lys Pro Thr Ser Pro Val Lys Pro Ile Pro Gln Asn Thr 1770 1775 1780	5 3 8 2
GAA TAT AGG ACA CGT GTA AGA AAA AAT GCA GAC TCA AAA AAT AAT TTA Glu Tyr Arg Thr Arg Val Arg Lys Asn Ala Asp Ser Lys Asn Asn Leu 1785 1790 1795	5430
AAT GCT GAG AGA GTT TTC TCA GAC AAC AAA GAT TCA AAG AAA CAG AAT Asn Ala Glu Arg Val Phe Ser Asp Asn Lys Asp Ser Lys Lys Gln Asn 1800 1805 1810 1815	5 4 7 8
TTG AAA AAT AAT TCC AAG GAC TTC AAT GAT AAG CTC CCA AAT AAT GAA Leu Lys Asn Asn Ser Lys Asp Phe Asn Asp Lys Leu Pro Asn Asn Glu 1820 1825 1830	5 5 2 6
GAT AGA GTC AGA GGA AGT TTT GCT TTT GAT TCA CCT CAT CAT TAC ACG Asp Arg Val Arg Gly Ser Phe Ala Phe Asp Ser Pro His His Tyr Thr 1835 1845	5 5 7 4
CCT ATT GAA GGA ACT CCT TAC TGT TTT TCA CGA AAT GAT TCT TTG AGT Pro Ile Glu Gly Thr Pro Tyr Cys Phe Ser Arg Asn Asp Ser Leu Ser 1850 1860	5622
TCT CTA GAT TTT GAT GAT GAT GAT GTT GAC CTT TCC AGG GAA AAG GCT Ser Leu Asp Phe Asp Asp Asp Asp Val Asp Leu Ser Arg Glu Lys Ala 1865 1870 1875	5670
GAA TTA AGA AAG GCA AAA GAA AAT AAG GAA TCA GAG GCT AAA GTT ACC Glu Leu Arg Lys Ala Lys Glu Asn Lys Glu Ser Glu Ala Lys Val Thr 1880 1885 1890	5718

45		46
	-continued	
Ser His Thr Glu Leu Th	C TCC AAC CAA CAA TCA GCT r Ser Asn Gln Gln Ser Ala	Asn Lys Thr Gin
	1905 A ATA AAT CGA GGT CAG CCC o lie Asn Arg Gly Gin Pro	
1915 CAG AAA CAA TCC ACT TT	1920 T CCC CAG TCA TCC AAA GAG e Pro Gln Ser Ser Lys Asj	1925 C ATA CCA GAC AGA 5862
1930	1935 A AAG TTA CAG AAT TTT GC	1940
	u Lys Leu Gln Asn Phe Ala 1950	a Ile Glu Asn Thr
Pro Val Cys Phe Ser Hi	T AAT TCC TCT CTG AGT TCC s As a Ser Ser Leu Ser Ses 1970	
	T AAA GAA AAT GAA CCT AT n Lys Glu Asn Glu Pro Il 1985	
	A GAA CCA AGT AAA CCT CA. y Glu Pro Ser Lys Pro Gl: 2000	
	T GTT GAA GAT ACC CCA GT s Val Glu Asp Thr Pro Va 2015	
	CT CTT AGT ATT GAC TCT GA. T Leu Ser Ile Asp Ser Gl 2030 20	n Asp Asp Leu Leu
Gin Glu Cys lle Ser Se	CC GCA ATG CCA AAA AAG AA FT Ala Met Pro Lys Lys Ly Lys Ly Ly S Ly S Ly S Ly S L	
	AA AAA CAT AGT CCC AGA AA u Lys His Ser Pro Arg As 2065	
	CA CTT GAT TTG AAA GAT AT ar Leu Asp Leu Lys Asp 11 2080	
	CC CCT GAT TCA GAA AAT TT or Pro Asp Ser Glu Asn Ph 2095	
	AT TCC ATA GTA AGT AGT TT sn Ser lle Val Ser Ser Le 2110 21	u His Gla Ala Ala
Ala Ala Ala Cys Leu Se	CT AGA CAA GCT TCG TCT GA er Arg Gln Ala Ser Ser As 125 2130	
	GAATC TCT CTG GGA TCA CC iy ile Ser Leu Giy Ser Pr 2145	
	AA CCC TTT ACA AGT AAT AA ys Pro Phe Thr Ser Asn Ly 2160	
	AA AGT ACA TTG GAA ACT AA ys Ser Thr Leu Glu Thr Ly 2175	
	AA GGA GGA AAA AAA GTT TA ys Gly Gly Lys Lys Val Ty 2190 21	
Thr Gly Lys Val Arg Sc	CT AAT TCA GAA ATT TCA GG er Asn Ser Glu Ile Ser Gl 205 2210	

-continued	
CCC CTT CAA GCA AAC ATG CCT TCA ATC TCT CGA GGC Pro Leu Gln Ala Asn Met Pro Ser Ile Ser Arg Gly 2220 2225	AGG ACA ATG ATT 6726 Arg Thr Met IIe 2230
CAT ATT CCA GGA GTT CGA AAT AGC TCC TCA AGT ACA His Ile Pro Gly Val Arg Asn Ser Ser Ser Thr 2235	
AAA AAA GGC CCA CCC CTT AAG ACT CCA GCC TCC AAA Lys Lys Gly Pro Pro Leu Lys Thr Pro Ala Ser Lys 2250	AGC CCT AGT GAA 6822 Ser Pro Ser Glu 2260
GGT CAA ACA GCC ACC ACT TCT CCT AGA GGA GCC AAG Gly Gln Thr Ala Thr Thr Ser Pro Arg Gly Ala Lys 2265 2270 2275	Pro Ser Val Lys
TCA GAA TTA AGC CCT GTT GCC AGG CAG ACA TCC CAA Ser Glu Leu Ser Pro Vai Ala Arg Gln Thr Ser Gln 2280 2285 2290	
AGT AAA GCA CCT TCT AGA TCA GGA TCT AGA GAT TCG Ser Lys Ala Pro Ser Arg Ser Gly Ser Arg Asp Ser 2300 2305	ACC CCT TCA AGA 6966 Thr Pro Ser Arg 2310
CCT GCC CAG CAA CCA TTA AGT AGA CCT ATA CAG TCT Pro Ala Gln Gln Pro Leu Ser Arg Pro Ile Gln Ser 2315	CCT GGC CGA AAC 7014 Pro Gly Arg Asn 2325
TCA ATT TCC CCT GGT AGA AAT GGA ATA AGT CCT CCT Ser lle Ser Pro Gly Arg Asn Gly lle Ser Pro Pro 2330	AAC AAA TTA TCT 7062 Asn Lys Leu Ser 2340
CAA CTT CCA AGG ACA TCA TCC CCT AGT ACT GCT TCA Gln Leu Pro Arg Thr Ser Ser Pro Ser Thr Ala Ser 2345 2350	Thr Lys Ser Ser
GGT TCT GGA AAA ATG TCA TAT ACA TCT CCA GGT AGA Gly Ser Gly Lys Met Ser Tyr Thr Ser Pro Gly Arg 2360 2370	Gln Met Ser Gln 2375
CAG AAC CTT ACC AAA CAA ACA GGT TTA TCC AAG AAT Gln Asn Leu Thr Lys Gln Thr Gly Leu Ser Lys Asn 2380	
CCA AGA AGT GAG TCT GCC TCC AAA GGA CTA AAT CAG Pro Arg Ser Glu Ser Ala Ser Lys Gly Leu Asn Gln 2395 2400	
AAT GGA GCC AAT AAA AAG GTA GAA CTT TCT AGA ATG Asn Gly Ala Asn Lys Lys Val Glu Leu Ser Arg Met 2410 2415	
TCA AGT GGA AGT GAA TCT GAT AGA TCA GAA AGA CCT Ser Ser Gly Ser Glu Ser Asp Arg Ser Glu Arg Pro 2425 2430 243	Val Leu Val Arg 5
CAG TCA ACT TTC ATC AAA GAA GCT CCA AGC CCA ACC Gln Ser Thr Phe Ile Lys Glu Ala Pro Ser Pro Thr 2440 2445	Leu Arg Arg Lys 2455
TTG GAG GAA TCT GCT TCA TTT GAA TCT CTT TCT CCA Leu Glu Glu Ser Ala Ser Phe Glu Ser Leu Ser Pro 2460 2465	Ser Ser Arg Pro 2470
GCT TCT CCC ACT AGG TCC CAG GCA CAA ACT CCA GTT Ala Ser Pro Thr Arg Ser Gln Ala Gln Thr Pro Val 2475 2480	Leu Ser Pro Ser 2485
CTT CCT GAT ATG TCT CTA TCC ACA CAT TCG TCT GTT Leu Pro Asp Met Ser Leu Ser Thr His Ser Ser Val 2490 2495	Gln Ala Gly Gly 2500
TGG CGA AAA CTC CCA CCT AAT CTC AGT CCC ACT ATA Trp Arg Lys Leu Pro Pro Asn Leu Ser Pro Thr Iie 2505 2510 251	Glu Tyr Asn Asp 5
GGA AGA CCA GCA AAG CGC CAT GAT ATT GCA CGG TCT Gly Arg Pro Ala Lys Arg His Asp Ile Ala Arg Ser 2520 2530	

47	7			50
		-continued		
CCT TCT AGA CTT CC	A ATC AAT AGG	TCA GGA ACC	TGG AAA CGT	GAG CAC 7686
Pro Ser Arg Leu Pro 25		Ser Gly Thr 2545		2550
23	40	2343		2330
AGC AAA CAT TCA TC	A TCC CTT CCT	CGA GTA AGC	ACT TGG AGA	AGA ACT 7734
Ser Lys His Ser Se				
2 5 5 5		2560	2 5 6 5	
GGA AGT TCA TCT TC.				
Gly Ser Ser Ser Se			Glu Ser Ser 2580	Glu Lys
2 5 7 0	2 5 7	5	2380	
GCA AAA AGT GAG GA	T GAA AAA CAT	GTG AAC TCT	ATT TCA GGA	ACC AAA 7830
Ala Lys Ser Glu As				
2 5 8 5	2590		2595	•
CAA AGT AAA GAA AA				
Gln Ser Lys Glu As				
2600	2605	2610	•	2 6 1 5
AAA GAA AAT GAA TT	T TOT CCC ACA	AAT AGT ACT	TOT CAG ACC	GTT TCC 7926
Lys Glu Asn Glu Ph				
26		2625		2630
-				
TCA GGT GCT ACA AA	T GGT GCT GAA	TCA AAG ACT	CTA ATT TAT	CAA ATG 7974
Ser Gly Ala Thr As	n Gly Ala Glu	Ser Lys Thr	Leu Ile Tyr	Gln Met
2635		2640	2645	
GCA CCT GCT GTT TC				
Ala Pro Ala Val Se				Glu Asp
2650	2 6 5	5	2660	
TGT CCC ATT AAC AA	T CCT AGA TCT	GGA AGA TCT	CCC ACA GGT	AAT ACT 8070
Cys Pro Ile Asn As				
2665	2670	0.,,	2675	
CCC CCG GTG ATT GA				
Pro Pro Val Ile As	p Ser Val Ser	Glu Lys Ala	Asn Pro Asn	Ile Lys
2680	2685	2690	0	2695
GAT TCA AAA GAT AA				
Asp Ser Lys Asp As	in Gin Ala Lys '00	Gin Ash Val	GIY ASE GIY	2710
2 1	00	2103		2710
CCC ATG CGT ACC GT	G GGT TTG GAA	AAT CGC CTG	ACC TCC TTT	ATT CAG 8214
Pro Met Arg Thr Va				
2715	•	2720	2725	
GTG GAT GCC CCT GA				
Val Asp Ala Pro As				Gln Asn
2730	2 7 3	5	2740	
AAT CCT GTC CCT GT		AAT GAA AGT	CCT ATA GTG	GAA CGT 8310
Asn Pro Val Pro Va				
2745	2750	014 501	2755	6
2,10				
ACC CCA TTC AGT TO	CT AGC AGC TCA	AGC AAA CAC	AGT TCA CCT	AGT GGG 8358
Thr Pro Phe Ser Se				
2760	2765	277	0	2775
ACT GTT GCT GCC AC				
Thr Val Ala Ala As		•	Asn Pro Ser	-
2 7	7 8 0	2785		2790
AAA AGC AGC GCA GA	AT AGC ACT TO	GCT CGG CCA	TCT CAG ATC	CCA ACT 8454
Lys Ser Ser Ala As				
2795	.,	2800	280	
CCA GTG AAT AAC AA	AC ACA AAG AA	G CGA GAT TCC	AAA ACT GAC	AGC ACA 8502
Pro Val Asn Asn A				Ser Thr
2810	2 8	1.5	2820	
			mam ass ===	m.a amm
GAA TCC AGT GGA AG Glu Ser Ser Gly Tl				
2825	nr Gin Ser Pr	, Lys AIG DIS	2835	Ly L L C G
2023	2030		2000	
GTG ACA TCT GTT TA	AAAGAGAG GAA	GAATGAA ACTAA	GAAAA TTCTAT	GTTA 8602
Val Thr Ser Val				
2840				

-continued

ATTACAACTG CTATATAGAC ATTTTGTTTC AAATGAAACT TTAAAAGACT GAAAAATTTT 8662 GTAAATAGGT TTGATTCTTG TTAGAGGGTT TTTGTTCTGG AAGCCATATT TGATAGTATA 8722 CTTTGTCTTC ACTGGTCTTA TTTTGGGAGG CACTCTTGAT GGTTAGGAAA AAATAGAAAG 8 7 8 2 CCAAGTATGT TTGTACAGTA TGTTTTACAT GTATTTAAAG TAGCATCCCA TCCCAACTTC 8842 CTTAATTATT GCTTGTCTAA AATAATGAAC ACTACAGATA GGAAATATGA TATATTGCTG 8902 TTATCAATCA TTTCTAGATT ATAAACTGAC TAAACTTACA TCAGGGGAAA ATTGGTATTT 8962 ATGCAAAAA AAAATGTTTT TGTCCTTGTG AGTCCATCTA ACATCATAAT TAATCATGTG 9022 9082 ACTGCATGAA TGAAACTGAT GGTTCAATTT CAGAAGTAAT GATTAACAGT TATGTGGTCA 9142 CATGATGTGC ATAGAGATAG CTACAGTGTA ATAATTTACA CTATTTTGTG CTCCAAACAA 9202 AACAAAAATC TGTGTAACTG TAAAACATTG AATGAAACTA TTTTACCTGA ACTAGATTTT 9262 ATCTGAAAGT AGGTAGAATT TTTGCTATGC TGTAATTTGT TGTATATTCT GGTATTTGAG 9322 GTGAGATGGC TGCTCTTTAT TAATGAGACA TGAATTGTGT CTCAACAGAA ACTAAATGAA 9382 CATTTCAGAA TAAATTATTG CTGTATGTAA ACTGTTACTG AAATTGGTAT TTGTTTGAAG 9442 GGTTTGTTTC ACATTTGTAT TAATTAATTG TTTAAAATGC CTCTTTTAAA AGCTTATATA 9502 AATTTTTTCT TCAGCTTCTA TGCATTAAGA GTAAAATTCC TCTTACTGTA ATAAAAACAT 9562 TGAAGAAGAC TGTTGCCACT TAACCATTCC ATGCGTTGGC ACTT 9606

(2) INFORMATION FOR SEQ ID NO:2:

æ

ļ.

H

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2843 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (i i) MOLECULE TYPE: protein
- ($\mathbf{x} \cdot \mathbf{i} \cdot$) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Me t	Ala	Ala	Ala	Ser 5	Туг	Asp	Gln	Leu	Leu 10	Lys	Gin	V a 1	G 1 u	Ala 15	Leu
Lys	Met	G 1 u	Asn 20	Ser	Asn	Leu	Arg	G 1 n 2 5	Glu	Leu	Glu	Asp	A s n 3 0	Ser	Asn
His	Leu	Thr 35	Lys	Leu	G1 u	Thr	G 1 u 4 0	Ala	Ser	Asn	Me t	L y s 4 5	G 1 u	V a 1	Leu
Lys	G 1 n 5 0	Leu	Gln	G 1 y	Ser	I 1 e 5 5	Glu	Аsр	G 1 u	A 1 a	Me t 6 0	Ala	Ser	Ser	Gly
G 1 n 6 5	I 1 e	Аsр	Leu	Leu	G 1 u 7 0	Arg	Leu	Lys	Glu	Leu 75	Asn	Leu	Asp	Ser	Ser 80
Asn	P h e	Pro	G 1 y	V a 1 8 5	Lуs	L e u	Агд	Ser	L y s 9 0	M e t	Ser	Leu	Arg	S e r 9 5	Туr
G 1 y	Ser	Агд	G 1 u 1 0 0	Gly	Ser	V a 1	Ser	Ser 105	Arg	Ser	G l y	Glu	C y s 1 1 0	Ser	Рго
V a 1	Pro	Me t	G 1 y	Ser	P h e	Pro	Arg 120	Агд	G 1 y	Phe	V a 1	A s n 1 2 5	Gly	Ser	Агд
Glu	S e r 1 3 0	Thr	G 1 y	Туг	Leu	G 1 u 1 3 5	Glu	Leu	Glu	Lys	G 1 u 1 4 0	Агд	Ser	Leu	Leu
Leu 145	A 1 a	Авр	Leu	Авр	L y s 1 5 0	Glu	Glu	Lys	Glu	Lys 155	Авр	Тгр	Туг	Туг	A 1 a 1 6 0
Gln	Leu	Gln	Asn	Le u 165	Thr	Lys	Агд	I 1 c	A s p 1 7 0	Ser	Leu	Pro	Leu	Thr 175	Glu

Asn Phe Ser Leu Gin Thr Asp Leu Thr Arg Arg Gin Leu Giu Tyr Giu

		23			54
			-contin	nued	
	1 8	0	1 8 5	1	190
Ala Arg	Gln Il 195	e Arg Val	Ala Met Glu G	iu Gln Leu Gly 7 205	Thr Cys Gln
Asp Met		s Arg Ala	Gln Arg Arg I	le Ala Arg Ile C 220	Gin Gin Ile
Glu Lys 225	ı Asp Il	e Leu Arg 230	Ile Arg Gln L	eu Leu Gln Ser (235	Gln Ala Thr 240
Glu Ala	a Glu Ar	g Ser Ser 245		is Glu Thr Gly 5	Ser His Asp 255
Ala Gla	Arg Gl 26		Gly Gln Gly V 265	al Gly Glu Ile A	Asn Met Ala 270
Thr Sea	Gly As 275	n Gly Gln	Giy Ser Thr Ti 280	hr Arg Met Asp I 285	His Glu Thr
Ala Sea 296		eu Ser Ser	Ser Ser Thr H 295	is Ser Ala Pro A	Arg Arg Leu
Thr Sea 3 0 5	r His Le	eu Gly Thr 310	Lys Val Glu M	et Val Tyr Ser 1 315	Leu Leu Ser 320
Met Le	ı Giy Ti	or His Asp 325		et Ser Arg Thr :	Leu Leu Ala 335
Met Se		er Gin Asp 40	Ser Cys Ile S 345	er Met Arg Gln	Ser Gly Cys 350
Leu Pr	0 Leu Le 355	eu lle Gln	Leu Leu His G 360	ly Asn Asp Lys 365	Asp Ser Val
Leu Le		sn Ser Arg	Gly Ser Lys G 375	lu Ala Arg Ala . 380	Arg Ala Ser
Ala Al 385	a Leu H	is Asn Ile 390	Ile His Ser G	ln Pro Asp Asp 395	Lys Arg Gly 400
-	-	405	4	eu Glu Gln Ile . 10	4 1 5
-	4	2 0	4 2 5		4 3 0
	4 3 5		4 4 0	al Glu His Gln 445	
4 5	0		4 5 5	he Asp Glu Glu 460	•
465		470	·	Ala Ile Ala Glu 475	480
		4 8 5	4	Asp His Tyr	4 9 5
		0 0	5 0 5	The Asn Leu The	5 1 0
	5 1 5		5 2 0	fet Lys Gly Cys 525	
5 3	0	·	5 3 5	Flu Asp Leu Gln 540	
5 4 5		5 5 6	_	555	5 6 0
-		5 6 5	5	Lys Ala Leu Met 570 Lys Ser Val Leu	5 7 5
	5	8 0	5 8 5	Lys Ser Val Leu Asn Lys Ala Asp	5 9 0
IIP A8	595	UI AIZ DI	600	ASH LYS AIR ASP 605	ii Cys Ala

				33						50
							-continued			
Val	A s p 6 1 0	G 1 y	Ala	Leu A	la Phe 615	Leu V	al Gly	Thr Leu 620	Thr Tyr	Arg Ser
G 1 n 6 2 5	Thr	Asn	Thr	Leu A	1 a I 1 e 3 0	Ile G		Gly Gly 635	Gly Ile	Leu Arg 640
Asn	Val	Ser	Ser	Leu I 645	le Ala	Thr A	sn Glu 650	Asp His	Arg Gln	11e Leu 655
Arg	Glu	Asn	Asn 660	Cys L	eu Gln		eu Leu 65	Gln His	Leu Lys 670	Ser His
Ser	Leu	Thr 675	I 1 e	Val S	er Asn	Ala C 680	ys Gly	Thr Leu	Trp Asn 685	Leu Ser
Ala	Arg 690	Asn	Pro	Lys A	sp Gln 695	Glu A	la Leu	Trp Asp 700	Met Gly	Ala Val
Ser 705	M e t	Lęu	Lys		eu Ile	His S	•	His Lys 715	Met Ile	Ala Met 720
G 1 y	Ser	Ala	Ala	Ala L 725	eu Arg	Asn L	eu Met 730	Ala Asn	Arg Pro	Ala Lys 735
Туг	Lys	Asp	A 1 a	Asn I	le Met		ro Gly 45	Ser Ser	Leu Pro 750	Ser Leu
His	V a 1	Arg 755	Lys	Gln L	.ys Ala	Leu G 760	lu Ala	Glu Leu	Asp Ala 765	Gln His
Leu	S c r 7 7 0	Glu	Thr	Phc A	Asp Asn 775	Ile A	sp Asn	Leu Ser 780	Pro Lys	Ala Ser
His 785	Arg	Ser	Lys		Arg His 790	Lys G	ln Ser	Leu Tyr 795	Gly Asp	Tyr Va1 800
Рbе	Asp	Thr	Asn	Arg H 805	His Asp	Asp A	s n Arg 810	Ser Asp	Asn Phe	Asn Thr 815
G 1 y	Asn	Met	Thr 820		Leu Ser		yr Leu 25	Asn Thr	Thr Val 830	Leu Pro
Ser	Ser	S e r 8 3 5		Ser A	Arg Gly	Ser L 840	си Азр	Ser Ser	Arg Ser 845	Glu Lys
Asp	Arg 850		Leu	Glu A	Arg Glu 855	Arg G	ly Ile	Gly Leu 860	Gly Asn	Tyr His
Pro 865		Thr	Glu		Pro Gly 870	Thr S	er Ser	Lys Arg 875	Gly Leu	Gln Ile 880
Ser	Thr	Thi	Ala	Ala (Gln Iie	Ala L	ys Vai 890	Met Glu	Glu Val	Ser Ala 895
Ile	His	Thi	S e 1		Glu Asp		er Ser	Gly Ser	Thr Thr 910	Glu Leu
His	Сув	Va 1		Asp (Glu Arg	Asn A 920	la Leu	Arg Arg	Ser Ser 925	Ala Ala
His	7 h r 9 3 0		S e :	r Asn '	Thr Tyr 935		he Thr	Lys Ser 940		Ser Asn
Arg 945		Су	S e :		Рго Туг 950	Ala I	ys Leu	Glu Tyr 955	Lys Arg	Ser Ser 960
Ası	ı Ası	Se	r Le	u Asn 965	Ser Val	Ser S	970	Asp Gly	Tyr Gly	Lys Arg 975
G 1 3	y Gli	м с	t Ly 98		Ser Ile		Ser Tyr 985	Ser Glu	Asp Asp 990	Glu Ser
Lys	s Pho	С у 9 9		r Tyr	Gly Gla	Tyr 1	Pro Ala	Asp Leu	Ala His 1005	Lyslie
Hi	8 Sen 10:		a As	n His	Met Ası 101		Asn Asp	Gly Glu 102		Thr Pro
I 1 6		n Ty	r Se		Lys Tyn 1030	Ser A	Asp Glu	Gln Leu 1035	Asn Ser	Gly Arg

																			_	α	nu	Ш	uc	u_																	
Gln	Ser	:	Рг	0	S	e I			n 4 5	A s	n	c) 1	u	A	r	g	T	'n	p			a 5 (r	g	P	r	0	L	y	s	н	i	s			e 5 5		c	
Glu	Asp	,	G I	u		l e) 6 (y	s	G 1	n	S	6 c	r	G	1	u		3 1 0			r	g	G	1	n	s	c	r	A	r	g			n 70		1	n	Se	; т	
Thr	Thr			r 75		ro	v	а	1	Т	r	3	Γh	r			u 8		; c	r	Τ	'n	r	A	s	p	A	s	p			s 8 5		i	s	L	c	u	Ly	7 S	
Phe	G 1 n			0	H	i s	P	h	c	G I	y		3 1 1 0			1	n	Ċ	7 1	u	c	у	s	v	a	1	S	e 1	r 0 (P 0	r	٥	T	y	r	A	r	g	S	e r	
Arg 1105		,	A 1	а	A :	s n	G	1		S 6			3 1	u	T	'n	r	A	l s	n	A	r	g	V 1	7 a	1	5	÷ 1	y	s	e	r	A	s	n	н	i	s		1 y 1 2	
I i e	Ası	1	G 1	n	A	s n			1 2 5		: 1	ď	3 1	n	s	c	r	I	. e	u			s 3		; 1	n	c	; 1	u	A	s	P	A	s	p			r 3 5		1 u	
Авр	Asj	p	Lу	s		r o 1 4		h	r	A :	s n	7	Гу	r	s	e	r		3 1 1 1			l r	g	7	у	r	5	c	r	c	; 1	u			u 5 (1	u	G	l n	
His	Glu			u 5 5	G			1	u	A:	r g	1	Рr	0			r 6	A				ſ y	r	s	6 e	r	1	1	c			8 6 :		у	ī	A	. s	n	G	l u	
Glu	L y :	s	Αг			i s	v	a	1	A	s p		G I 1 1		P				I 1	e	A	l s	p	1	Гy	r			r 8	I				y	8	T	y	r	A	l a	
Thr	Asj			c	P	ro	s	c	r		e r 19	•			-	. y	8	•	3 1	n	8	S c	r			c					• h	c	s	c	r	L	y	s		e r 2 0	
Ser		r	G 1	y	G	1 n			r 0 5	s			Lу	s	7	: h	r	(3 1	u			s 2 1	N	_	_	-	S c	ī	S	6 e	r	s	c	r			r 1 :	G		
Asn	T b	T	Se	ŧ			I				е г		Se	r	A	l s	n				1				٩r	g	•	3 1	n	4	l s	n				L				i s	ı
Pro	S e	r			A	22 1a		3 1	n	s	e r		Αr	g				•	1 2 G 1			3 1	l n	1	Pr	0	•	3 1	n				A		3 (a		. 1	a	T	h r	•
Сув			V s	. 3 5		e r		S e	r	I	l c				(. 4		G I	u		Γŀ	1 1		I 1	. e				:		4 1 T		` y	r	c	y	8	v	a l	l
Glu	12 As			1 T	P	ro	1	1 1	e	С	y s		1 2 P b			5 6	r		A 1	g	: (c 5	y s	;	Se	: 1			: 6 : r		i. e	e u	s		r	s	i c	r	L	e t	1
126 Ser		ī	A !	l a	G	1 u		A s	p		27 1 u		1 1	l e	(3 1	و ا		c s	7 8		A. s	s n			2 7 1 n		ГĦ	1 1	•	ri	1 1	(; 1	n	c	; 1	u			3 0
Asp	Se	ī	A	la	A	s n			8 5		e u	ı	G I	l na		1 1	le		A I	l a			29 1 u		11	ı e		Ŀ	7 8		3 1	ίν	I	. ¥	. 8			9 :		1 3	,
Thr					1	3 0	0												1 3	3 (5												1	1 3	1	0					
			1 :	3 1	5						_					1 3	3 2	0													1 3	3 2	5								
His	1 3	3 ()										1 3	3 3	5													1 :	3 4	0											
Glu 134	5					Ī				1	3 :	5 0													1 :	3 5	5												1	3 (6 0
Pro								1 3	6 :	5												1	3 7	0													13	7	5	-	
Val					1	3 8	0												1	3 1	B 5													1 3	3 9	0					
Leu	A s	P		с г 3 9		h e	•	G I	u	S	e :	r	A	r g				0		1 4	•	A	1 a	ı	S	e I		S	СГ			a 1 4 0		3 1	l n	;	S	r	G	1 1	u

Pro Cys Ser Gly Met Val Ser Gly Ile Ile Ser Pro Ser Asp Leu Pro 1410

Asp Ser Pro Gly Gln Thr Met Pro Pro Ser Arg Ser Lys Thr Pro Pro 1425

Pro Pro Pro Gln Thr Ala Gln Thr Lys Arg Glu Val Pro Lys Asn Lys 1455

Ala Pro Thr Ala Glu Lys Arg Glu Ser Gly Pro Lys Gln Ala Ala Val

- argan

	-continued	
1 4 6 0	1 4 6 5	1 4 7 0
Asn Ala Ala Val Gln Ar	g Val Gln Val Leu Pro Asp A	la Asp Thr Leu
1475	1480 1	485
Leu His Phe Ala Thr Gl 1490	u Ser Thr Pro Asp Gly Phe S 1495 1500	er Cys Ser Ser
	r Leu Asp Glu Pro Phe Ile G 10 1515	ln Lys Asp Val 1520
Glu Leu Arg Ile Met Pr	o Pro Val Gln Glu Asn Asp A	sn Gly Asn Glu
1525	1530	1535
Thr Glu Ser Glu Gln Pr	o Lys Glu Ser Asn Glu Asn G	ln Glu Lys Glu
1540	1545	1550
Ala Glu Lys Thr Ile As	p Ser Glu Lys Asp Leu Leu A	sp Asp Ser Asp
1555	1560 1	565
Asp Asp Asp Ile Glu II	e Leu Glu Glu Cys Ile Ile S 1575 1580	er Ala Met Pro
	rs Gly Lys Lys Pro Ala Gin T 590 1595	hr Ala Ser Lys 1600
Leu Pro Pro Pro Val Al	la Arg Lys Pro Ser Gln Leu P	ro Val Tyr Lys
1605	1610	1615
Leu Leu Pro Ser Gln As	sn Arg Leu Gln Pro Gln Lys H	is Val Ser Phe
1620	1625	1630
Thr Pro Gly Asp Asp Me	et Pro Arg Val Tyr Cys Val G	lu Gly Thr Pro
1635	1640 1	645
Ile Asn Phe Ser Thr Al	la Thr Ser Leu Ser Asp Leu T 1655 1660	hr Ile Glu Ser
	la Ala Gly Glu Gly Val Arg G 670 1675	Hy Gly Ala Gln 1680
Ser Gly Glu Phe Glu Ly	ys Arg Asp Thr Ile Pro Thr G	ilu Gly Arg Ser
1685	1690	1695
Thr Asp Glu Ala Gln Gl	ly Gly Lys Thr Ser Ser Val T	Thr Ile Pro Glu
1700	1705	1710
Leu Asp Asp Asn Lys Al	la Glu Glu Gly Asp Ile Leu A	ala Glu Cys Ile
1715	1720 1	1725
Asn Ser Ala Met Pro Ly 1730	ys Gly Lys Ser His Lys Pro P 1735 1740	he Arg Val Lys
Lys Ile Met Asp Gln Va	al Gln Gln Ala Ser Ala Ser S	Ser Ser Ala Pro
1745 12	750 1755	1760
Asn Lys Asn Gln Leu As	sp Gly Lys Lys Lys Lys Pro 1	Thr Ser Pro Val
1765	1770	1775
Lys Pro Ile Pro Gln As	sn Thr Glu Tyr Arg Thr Arg V	Val Arg Lys Asn
1780	1785	1790
Ala Asp Ser Lys Asn A:	sn Leu Asn Ala Glu Arg Val I	Phe Ser Asp Asn
1795	1800	1805
Lys Asp Ser Lys Lys G: 1810	ln Asn Leu Lys Asn Asn Ser I 1815 1820	ys Asp Phe Asn
	sn Glu Asp Arg Val Arg Gly S 830 1835	Ser Phe Ala Phe 1840
Asp Ser Pro His His T	yr Thr Pro Ile Glu Gly Thr I	Pro Tyr Cys Phe
1845	1850	1855
Ser Arg Asn Asp Ser L 1860	eu Ser Ser Leu Asp Phe Asp 1865	Asp Asp Asp Val 1870
Asp Leu Ser Arg Glu L	ys Ala Glu Leu Arg Lys Ala 1	Lys Glu Asn Lys
1875	1880	1885

_																									COI	ш	шu	cu																		
7	G 1	u		; r 3 9		, 1	u	A	. 1	а	L	у	s	V	а	1			r 9 5		e :	r	н	i	s	Т	h 1	r	G	1 ι			e u		ГЬ	r	s	c	ı	A	S	Ω	G	1 1	D	
	3-1 1-9			r	A	. 1	a	A	. 8	Ω	L	y	s			r 10		1 :	n	A	1 :	a	I	1	c	A	1 :				s 1 5		l n		Рг	0	I	1	e	A	s :	n			g 2 0	
,	G 1	y	G I	l n	1	r	0	L	y	s			o 2 :		1	e	L	c	u	G	1 :	n	L	y	s			n 3 O		e :	r	Т	Ьr		Рb	c	P	' r	0			n 3 5		e :	r	
	Se	ī	Ly	, s	Å	à s	p			c 4		r	0	A	. s	p	A	r	g	G	1	y			a 4 5		1 :	a	T	h :	r	A	s p		G 1	u			s 5 0		e	u	G	1	n	
	A s	n	ΡI	h c		1 1 9			1	¢	G	, 1	u	A	s	n	T	h	r			o 6 0		a	1	С	y	8	P	h ·	e	s	ег			. s		L S	n	s	c	r	s	c	ī	
	Le	u		e r 9 7		S e	r	L	. е	u	s	c	r	A	. 8	P			e 7 5		s	P	G	1	n	G	1	u	A	s :			s n 98		A s	n	L	. y	s	G	1	u	A	s	n	
	G I 1 9			ro	:	1	c	L	, у	s	c	7 1	u			r 9 (1	u	P	r	0	P	r	o	A	S	P			r 9 5		l n		G I	y	C	; 1	u	P	r	0			r 0 0	ı
	Lу	s	P	ro	•	3 1	n	A	. 1	а			r O		; 1	y	Т	y	r	A	1	a	P	r	0			s 1 0		c	r	P	b e		Ηi	S	١	7 a	1			u 15		. 8	p	
	ТЬ	r	P	ro	,	Vа	. 1			s 2		h	c	S	c	r	A	r	g	A	s	n			r 2 5		e	r	L	e	u	s	e r		Se	; r			u 3 (c	r	I	1	c	
	A s	P	s	e r		G 1			l s	P	I	l s	p	I	. е	u	L	c	u			n 40		1	u	С	y	s	I	1	e	s	e r			e r		۱ 1	а	М	[e	t	P	ī	0	
	Lу	s		y s 0 5		•	s	I	. y	8	1	? 1	0	5	6 e	r			g 5 5		e	u	L	y	s	G	1	y	A	s	P		s 1		G I	lu	. 1	. у	· s	H	i	s	S	e	T	
	P r 2 0			гg		A. s	n	N	A e	t	(3 1	y			у 7 (1	c	L	c	u	G	1	y	G	1	u			P 7 5		c t		Tŀ	ı	1	<u>.</u> c	11	A	S	P			u 80)
	Lу	8	A	s p		1 1	c	(3 1	n			g 8		r	٥	A	. s	P	s	c	r	G	1	u			s 9 (1	y	L	c ı	1	S	• 1	1	? r	0			P 9 :		e	r	
	G 1	u	A	s n	ı	Ρł	ı c			р 0		Гг	p	1	. у	8	A	. 1	a	I	1	c			n 0 :		1	u	G	1	y	A	1 2	1	A:	5 B			; r		1	c	v	r a	I	
	S e	r	s	e I		L 6			Ηi	8	•	3 1	n	1	1	a	A	. 1	a			a 2 C		. 1	a	A	. 1	a	С	y	8	L	e i	1			2 5	A r	g	G	} 1	n	A	. 1	a	
	S c	r		e r 1 3			p	;	Se	r		A s	p	:	S e	r			e 3 :		e	u	S	e	r	L	e	u	L	y	8		e 1			l y	, ;	1 1	e	s	; e	r	L	. c	u	
	G 1 2 1			e I	•	P:	. 0	1	P 1	1 6	1	H i	. \$			u 5		ъ	r	P	ſ	0	A	L S	р	G	1	Ð			u 5 5		1 1	2	L	y s	.]	Pı	: 0	P	b	e			r 6 0	D
	S e	r	A	8 1	1	L;	ys	•	G I	l y			6		A r	g	I	1	e	L	. c	u	1	. у	g			o 7 (1	у	G	1 1	1	L;	y s		Se	e r			7		. c	u	
	G 1	u	T	ь	•	L ;	y s	:	L : 2 :	y s 18	0	1 1	c	,	G 1	u	S		ſ	G	i 1	u	2	S e	г 8	1 5	, у	8	G	1	y	I	1	c	L	y s	; (G 1	l y 19	0	3 1	y	I	_ y	s	
	Ly	s	V	a I				5		y s		S	r		Le	u	1	1				r 0 (3 1	y	I	. y	g	٧	a	1	A	r	g			0.5		3 n	S	3 c	T	C	3 1	u	
	I 1	c		e 1 2 1			1 y		G	l n	ι :	M (ŧ		L 3				n 1		r	0	1		e u	•	7 1	n	A	. 1	a		. 8			e t	t	P 1	ro	S	S e	: 1	1	1	c	
		r 2 2		. г į	g	G	1 у	•	A	rg	;	T 1	1 F			t 2 3		1	c	E	Ιi	8	1	[]	l c	I	, 1	٥			у 3 :		a	1	A	rg	3	A :	s n	8	S e	: I			: r	
	S	; r	S	e :	r	T	b r		S	e r			2 4			1	8	S e	T	I	. y	g	I	L 3	7 8			y 5		r	0	P	ī	0	L	e t	a	L ;	y s			1 T 2 5			0	
~	A :	l a	s	c	r	L	y s				: 5 0		ro		s	r	(3 1	u	(3 1	y			i n 26		ГЪ	r	A	. 1	a	7	h	r	T	h 1			е г 2 7		Pτ	0	1	A r	g	
	G :	l y	A	.1				3 7 5		r	•	s	е г		V :	a 1	1	L 3	7 8			: r		G :	lu	1	i e	u	8	S e	r	I	r				1 8 5		1 a	4	A I	g	(3 I	n	
	T	hт		e 2			1 1	1	I	1 6	•	G	1 y	,	G	Ìу			; r 2 9		S 6	r	1	L	y s		4. 1	. a	I	? 1	0		3 2 3			rį	g	s	e r	(G I	l y	;	Sé	; r	
		r g 3 0		8	p	s	e i	r	T	h i	r	P	rc			e r 3 1			g	1	Pı	. 0		A.:	l a	•	3 1	n			. n		r	0	L	c i	u	s	ег		A 1	r g			7 o	

2 3 5 5

2365

Ile Gin Ser Pro Gly Arg Asn Ser Ile Ser Pro Gly Arg Asn Gly Ile 2325 2330 2335

Ser Pro Pro Asn Lys Leu Ser Gin Leu Pro Arg Thr Ser Ser Pro Ser

Thr Ala Ser Thr Lys Ser Ser Gly Ser Gly Lys Met Ser Tyr Thr Ser

2360

Pro		G 1				r	g	•	G 1	n	;	M	c 1	t	s	c	1		3 1				i r	1	A:	S 1	1	L	u		T ł	ır				3 3 0		1 :	n	Т	h	r	G) 1	у]	<u>.</u>
S e r 2 3 8			y :	3	A	s	n		A. 1	. а		S	e 1	r			r 9 (1	c		P:	rc	•	A :	ΓĘ	3	S	r		G 1	u 3 9	5	s	C 1	r	A	1 :	a	s	c	r	L	. y	s		G :
Leu		A :	S 1	n.	G	1	n	1	M d	; t		A 2	s :	n 0 5	A	s	n	(3 1	у		A	S 1	n	G	1 ,	y	A :	1 a 4 1	0	A	s n		L :	y :	s	L	y	s	v	a	1	2	3 1 2 4	ี บ • 1	5	L
Ser		A 1	. 1	g	М	e	t				0		c :	r	T	h	r	Ι	_ y	s		s	C 1				r 2 5		lу		s	ı		G	1 1	1	s	e	ī			р 3 (l r	g	,	s
Glu	ı	A :	T (g				5	V a	ıl		L	c i	u	v	a	I	A	٩r	g				n 4 0		e 1	r	T !	h r		Ρĵ	h c		I	1 4	е			s 4 :		i 1	u	A	\ 1	. а		P
Ser				o 5 C		ь	r		L «	u		A	r	g	A	r	g		L y 2 4				c 1	a	G	1 1	n	G	l u		s	e r				a 6 0		c	r	P	h	c	c	3 1	u		S
Leu 246			e :	r	P	F	0		s	. 1		s	e	ī			g 7		Pτ	0		A	1 :	a	s	¢:	r	P	гo		T 1				rı	g	s	e	r	G	i 1	n	A	4 I	a		G 2
Thr	•	P	r	0	v	a	. 1		L	e t	ı			r 8 :		r	0	;	Se	r		L	c 1	u	P	r	0		s p			e t		s	e	r	L	е	u	s	e e	r			1 T 4 9		
Ser	•	s	e	ī	٧	a	. 1				ı) 0		1	a	G	1	y	•	G 1	l y		T	r j	p			g 0 5		y s		L	e u	ı	P	r	0	P	r	0			n 1		L c	; u		5
Рго	•	T	h	r				5		1 ι	ı	Т	у	ī	A	. s	n		A. s	s p	,			y 2 0		r	g	P	1 0	•	A	l a	ı	L	y	s			g 2		Ιi	8	I	A s	ı p		
Ala	1			g 3 (; e	r		H	i		s	¢	r	G	1	u		S 6				r	0	s	c	r	A	r g	ţ	L	e u	ı			o 4 (1	c	A	l s	n	A	A z	r g		:
G 1 y 2 5 4			h	r	7	? x	p	,	L	y i	3	A	r	g			12 5			is		s	e	r	L	y	8	H	is	1		e 1 5 5			c	r	s	c	r	I	. e	u	1	Pı	г о	ı	4
V a 1	ı	s	e	r	7	r 1t	r		T	r j	,			g 6 :		r	g		T I	h 1		G	i	y	s	¢	r		6 1 5 1			C 1	•	s	c	r	1	i	e	I	L e	u			e r 5 7		
Ser	r	s	e	r	(3 1	u	ı			B 0		e	r	G	1	u		L :	y s	ı	A	1	a			s 8 5		C 1	r	G	1 τ	1	A	8	p	G	1	u			8			i s	ı	
Asn	n	s	c	r				5		e :	r	G	1	y	7	ъ	T		L;	y	\$			n 0 C		c	T	L	y i	3	G	1 1	1	A	8	n			n 0		V a	1	;	s	e r	•	
Lys	s			y 1		Г 1	3 1		T	r	P	A	r	g	I	. y	7 8		I 2				y	s	G	1	u	A	8 1	Q	G	1 1	1			e 2		e	r	1	Pι	0	•	T	Ьr	:	
S e r 2 6 2			b	r	\$	s (6 1		G	1	R	T	h	r			1 5 3			c :	r	s	c	r	G	1	y	A	1 :	a		h 1			s	n	G	† 1	y	1	A 1	a	,	G	1 υ	1	
Lуs	s	T	h	T	1	L (e t	1	1	1	e			1 4		3 1	l n		М	e	t	A	1	a	P	r	0		6			a :	i	S	e	r	I	, у	8	7	Τł	1 T			1 u 6 s		
Vai	1	T	r	p	•	v :	a 1	I			g 6 (1	e	(3 1	lu		A	s :	p	c	y	5			o 6 5		1	е	A	8 :	n	A	. 8	n	F	• 1	0				0		e 1	r	
Arg	g	s	e	r				o 7 5		b	ī	C	1	y	4	A. :	s n	L	T	h	ī			o 8 (r	0	V	a	1	I	1	е	A	. 8	P				5		ı 1		s	e 1	c	
Lys	8			a 9		A	S 1	n.	P	r	0	A	s	n		1	l e	:		у б			. 8	P	s	e	r	L	y	В	A	8	P			n 0		3 1	n		A :	la		L	y :	5	
A s 1			a	. 1	•	G	1 :	y	A	5	n	C	} 1		;					a	1	P	r	0	M	í e	t	A		g					a	1	(3 1	y	:	L	e u		G	1 1	u	

Arg Leu Thr Ser Phe Ile Gin Val Asp Ala Pro Asp Gln Lys Gly Thr

Glu Ile Lys Pro Gly Gln Asn Asn Pro Val Pro Val Ser Glu Thr Asn

that the major to the same that the same tha H.

		-continued	
	2740	2745	2750
Glu	Ser Pro 11e Val 6	Glu Arg Thr Pro Phe Ser S 2760	er Ser Ser Ser Ser 2765
Lys	His Ser Ser Pro	Ser Gly Thr Val Ala Ala A	rg Val Thr Pro Phe
	2770	2775 2	780
A s n		Pro Arg Lys Ser Ser Ala A	sp Ser Thr Ser Ala
2 7 8		2790 2795	2800
Arg	Pro Ser Gln Ile	Pro Thr Pro Val Asn Asn A	sn Thr Lys Lys Arg
	2805	2810	2815
Азр	Ser Lys Thr Asp	Ser Thr Glu Ser Ser Gly T	hr Gln Ser Pro Lys
	2820	2825	2830
Агд	His Ser Gly Ser	Tyr Leu Val Thr Ser Val	

(2) INFORMATION FOR SEQ ID NO:3:

2835

- ($\,i\,$) SEQUENCE CHARACTERISTICS:

 - (A) LENGTH: 3172 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
- (i i) MOLECULE TYPE: cDNA
- (v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens
- (v i i) IMMEDIATE SOURCE:
 - (B) CLONE: DP1(TB2)
- (i \boldsymbol{x}) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..630

($\mathbf{x}\ \mathbf{i}\)$ SEQUENCE DESCRIPTION: SEQ ID NO:3:

														GGN Gly		4 8
1				5					10					1 5		
GAG	ACG	GTC	ccc	GCC	ATG	TCT	GCG	GCC	ATG	AGG	GAG	AGG	TTC	GAC	CGG	9 6
			Pro					Ala						Аsр		
			20					2 5					3 0			
TTC	CTG	CAC	GAG	AAG	AAC	TGC	ATG	ACT	GAC	CTT	CTG	GCC	AAG	CTC	GAG	1 4 4
Phe	Leu		Glu	Lys	Asn	Сув		Thr	Аsр	Leu	Leu		Lys	Leu	Glu	
		3 5					4 0					4 5				
GCC	AAA	ACC	GGC	GTG	AAC	AGG	AGC	TTC	ATC	GCT	CTT	GGT	GTC	ATC	GGA	192
Ala		Thr	G 1 y	Val	Asn		Ser	Рhе	Ιlε	Ala		Gly	V a 1	Ile	G 1 y	
	5 0					5 5					60					
CTG	GTG	GCC	TTG	TAC	CTG	GTG	TTC	GGT	TAT	GGA	GCC	TCT	CTC	CTC	TGC	2 4 0
	V a 1	Ala	Leu	Туг		V a 1	Phe	Gly	Туг	•	Ala	Ser	Leu	Leu	•	
6 5					70					7 5					8 0	
AAC	CTG	ATA	GGA	TTT	GGC	TAC	CCA	GCC	TAC	ATC	TCA	ATT	AAA	GCT	ATA	288
Asn	Leu	I 1 e	Gly		Gly	Туг	Pro	Ala	•	I i e	Ser	I 1 e	Lys	Ala	Ile	
				8 5					90					9 5		
GAG	AGT	ccc	AAC	AAA	GAA	GAT	GAT	ACC	CAG	TGG	CTG	ACC	TAC	TGG	GTA	3 3 6
Glu	Ser	Pro		Lys	Glu	Авр	Авр		Gln	Trp	Leu	Thr	•	Тrр	V a 1	
			100					105					110			
GTG	TAT	GGT	GTG	TTC	AGC	ATT	GCT	GAA	TTC	TTC	TCT	GAT	ATC	TTC	CTG	3 8 4
V a 1	Туг	-	Va1	Phe	Ser	I 1 e		Glu	Phe	Рhе	Ser	-	I 1 c	Phe	Leu	
		1 1 5			٠		120					1 2 5				
TCA	TGG	TTC	ccc	TTC	TAC	TAC	ATG	CTG	AAG	TGT	GGC	TTC	CTG	TTG	TGG	4 3 2
Ser	-	Phe	Pro	Phe	Тут		Met	Leu	Lys	Cys	•	Рье	Leu	Leu	Тrр	
	1 3 0					1 3 5					1 4 0					
TGC	ATG	GCC	CCG	AGC	CCT	TCT	AAT	GGG	GCT	GAA	CTG	CTC	TAC	AAG	CGC	480

-continued	
Cys Met Ala Pro Ser Pro Ser Asn Gly Ala Glu Leu Leu Tyr Lys Arg 145 150 155 160	
ATC ATC CGT CCT TTC TTC CTG AAG CAC GAG TCC CAG ATG GAC AGT GTG Ile Ile Arg Pro Phe Phe Leu Lys His Glu Ser Gln Met Asp Ser Val 165 170 175	5 2 8
GTC AAG GAC CTT AAA GAC AAG TCC AAA GAG ACT GCA GAT GCC ATC ACT Val Lys Asp Leu Lys Asp Lys Ser Lys Glu Thr Ala Asp Ala Ile Thr 180 185	5 7 6
AAA GAA GCG AAG AAA GCT ACC GTG AAT TTA CTG GGT GAA GAA AAG AAG Lys Glu Ala Lys Lys Ala Thr Val Asn Leu Leu Gly Glu Glu Lys Lys 195 200 205	6 2 4
AGC ACC TAAACCAGAC TAAACCAGAC TGGATGGAAA CTTCCTGCCC TCTCTGTACC Ser Thr 210	680
TTCCTACTGG AGCTTGATGT TATATTAGGG ACTGTGGTAT AATTATTTTA ATAATGTTGC	7 4 0
CTTGGAAACA TTTTTGAGAT ATTAAAGATT GGAATGTGTT GTAAGTTTCT TTGCTTACTT	800
TTACTGTCTA TATATATAGG GAGCACTTTA AACTTAATGC AGTGGGCAGT GTCCACGTTT	860
TTGGAAAATG TATTTTGCCT CTGGGTAGGA AAAGATGTAT GTTGCTATCC TGCAGGAAAT	920
ATAAACTTAA AATAAAATTA TATACCCCAC AGGCTGTGTA CTTTACTGGG CTCTCCCTGC	980
ACGSATTTC TCTGTAGTTA CATTTAGGRT AATCTTTATG GTTCTACTTC CTRTAATGTA	1040
CAATTTTATA TAATTCNGRA ATGTTTTTAA TGTATTTGTG CACATGTACA TATGGAAATG	1100
TTACTGTCTG ACTACANCAT GCATCATGCT CATGGGGAGG GAGCAGGGGA AGGTTGTATG	1 1 6 0
TGTCATTTAT AACTTCTGTA CAGTAAGACC ACCTGCCAAA AGCTGGAGGA ACCATTGTGC	1 2 2 0
TGGTGTGGTC TACTAAATAA TACTTTAGGA AATACGTGAT TAATATGCAA GTGAACAAAG	1 2 8 0
TGAGAAATGA AATCGAATGG AGATTGGCCT GGTTGTTTCC GTAGTATATG GCATATGAAT	1340
ACCAGGATAG CTTTATAAAG CAGTTAGTTA GTTAGTTACT CACTCTAGTG ATAAATCGGG	1 4 0 0
AAATTTACAC ACACACACA ACACACACA ACACACAC ACACACACA ACACACACA	1460
AGTACCCTGT AACTCTCAAT TCCCTGAAAA ACTAGTAATA CTGTCTTATC TGCTATAAAC	1520
TTTACATATT TGTCTATTGT CAAGATGCTA CANTGGAMNC CATTTCTGGT TTTATCTTCA	1580
NAGSGGAGAN ACATGTTGAT TTAGTCTTCT TTCCCAATCT TCTTTTTTAA MCCAGTTTNA	1640
GGMNCTTCTG RAGATTTG YC CACCTCTGAT TACATGTATG TTCT YGTTTG TATCATKAGC	1700
AACAACATGC TAATGRCGAC ACCTAGCTCT RAGMGCAATT CTGGGAGANT GARAGGNWGT	1760
ATARAGIMNC CCATAATCIG CITGGCAATA GITAAGICAA TCTATCITCA GITTITCTCT	1820
GGCCTTTAAG GTCAAACACA AGAGGCTTCC CTAGTTTACA AGTCAGAGTC ACTTGTAGTC	1880
CATTTAAATG CCCTCATCCG TATTCTTTGT GTTGATAAGC TGCACAKGAC TACATAGTAA	1940
GTACAGANCA GTAAAGTTAA NNCGGATGTC TCCATTGATC TGCCAANTCG NTATAGAGAG	2000
CAATTTGTCT GGACTAGAAA ATCTGAGTTT TACACCATAC TGTTAAGAGT CCTTTTGAAT	2060
TAAACTAGAC TAAAACAAGT GTATAACTAA ACTAACAAGA TTAAATATCC AGCCAGTACA	2120
GTATTTTTA AGGCAAATAA AGATGATTAG CTCACCTTGA GNTAACAATC AGGTAAGATC	2 1 8 0
ATNACAATGT CTCATGATGT NAANAATATT AAAGATATCA ATACTAAGTG ACAGTATCAC	2 2 4 0
NNCTAATATA ATATGGATCA GAGCATTTAT TTTGGGGAGG AAAACAGTGG TGATTACCGG	2300
CATTTTATTA AACTTAAAAC TTTGTAGAAA GCAAACAAAA TTGTTCTTGG GAGAAAATCA	
ACTITIAGAT TAAAAAAATT TTAAGTAWCT AGGAGTATTT AAATCCTTTT CCCATAAATA	2 4 2 0
AAAGTACAGT TTTCTTGGTG GCAGAATGAA AATCAGCAAC NTCTAGCATA TAGACTATAT	2 4 8 0
AATCAGATTG ACAGCATATA GAATATATTA TCAGACAAGA TGAGGAGGTA CAAAAGTTAC	2540
	•

TATTGCTCAT AATGACTTAC AGGCTAAAAN TAGNTNTAAA ATACTATATT AAATTCTGAA 2600 TGCAATTTTT TTTTGTTCCC TTGAGACCAA AATTTAAGTT AACTGTTGCT GGCAGTCTAA 2660 GTGTAAATGT TAACAGCAGG AGAAGTTAAG AATTGAGCAG TTCTGTTGCA TGATTTCCCA 2720 AATGAAATAC TGCCTTGGCT AGAGTTTGAA AAACTAATTG AGCCTGTGCC TGGCTAGAAA 2780 ACAAGCGTTT ATTTGAATGT GAATAGTGTT TCAAAGGTAT GTAGTTACAG AATTCCTACC 2840 AAACAGCTTA AATTCTTCAA GAAAGAATTC CTGCAGCAGT TATTCCCTTA CCTGAAGGCT 2900 TCAATCATTT GGATCAACAA CTGCTACTCT CGGGAAGACT CCTCTACTCA CAGCTGAAGA 2960 AAATGAGCAC ACCCTTCACA CTGTTATCAC CTATCCTGAA GATGTGATAC ACTGAATGGA 3020 AATAAATAGA TGTAAATAAA ATTGAGWTCT CATTTAAAAA AAACCATGTG CCCAATGGGA 3080 AAATGACCTC ATGTTGTGGT TTAAACAGCA ACTGCACCCA CTAGCACAGC CCATTGAGCT 3 1 4 0 ANCCTATATA TACATCTCTG TCAGTGCCCC TC 3172

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 210 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(i i) MOLECULE TYPE: protein

(x i) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Ala 1	V a 1	Ala Al	a Pro Val	Tyr Pro	Ala Leu 10	Gly Thr	Ala Pro	Gly Gly 15
Glu	Thr		o Ala Met	Ser Ala	Ala Met 25	Arg Glu	Arg Phe 30	Asp Arg
Phe	Leu	His G1	u Lys Asn	Cys Met 40	Thr Asp	Leu Leu	Ala Lys 45	Leu Glu
Ala	L y s 50	Thr G1	y Val Ass	Arg Ser 55	Phe Ile	Ala Leu 60	Gly Val	Ile Gly
Leu 65	V a 1	Ala Le	u Tyr Let 7(Gly Tyr	Gly Ala 75	Ser Leu	Leu Cys 80
Asn	Leu	Ile Gl	y Phe Gly 85	Tyr Pro	Ala Tyr 90	Ile Ser	Ile Lys	Ala Ile 95
Glu	Ser	Pro As	in Lys Gla)0	. Азр Азр	Thr Gln 105	Trp Leu	Thr Tyr 110	Trp Val
Val	Туг	Gly Va 115	al Phe Sen	1 1 e A 1 a 1 2 0		Phe Ser	Asp I1e 125	Phe Leu
Ser	Trp 130	Phe Pr	o Phe Ty	Tyr Met 135	Leu Lys	Cys G1y 140	Phe Leu	Leu Trp
Cys 145	Met	Ala Pi	ro Ser Pro		Gly Ala	Glu Leu 155	Leu Tyr	Lys Arg 160
I 1 e	Ile	Arg Pi	ro Phe Phe 165	e Leu Lys	His Glu 170	Ser Gln	Met Asp	Ser Val
Val	Lys		eu Lys Asj 80	Lys Ser	Lys Glu 185	Thr Ala	Asp Ala 190	Ile Thr
Lys	Glu	Ala Ly 195	ys Lys Ala	a Thr Val 200		Leu Gly	Glu Glu 205	Lys Lys
Ser	Thr							

```
( i ) SEQUENCE CHARACTERISTICS:
    ( A ) LENGTH: 434 amino acids
    ( B ) TYPE: amino acid
    ( C ) STRANDEDNESS: single
    ( D ) TOPOLOGY: linear
```

- (1 i) MOLECULE TYPE: protein
- ($v\ i\)$ ORIGINAL SOURCE: ($A\)$ ORGANISM: Homo sapiens

(v i i) IMMEDIATE SOURCE: (B) CLONE: TB1

(\mathbf{x} i) SEQUENCE DESCRIPTION: SEQ ID NO:5:

V a 1 1	Ala	Pro	V a 1	Val V	Val Gl	y	Ser	G l y	Arg 10	Ala	Pro	Агд	His	Pro 15	A 1 a
Рго	Ala	Ala	M c t 20	His 1	Pro An	g	Агд	Pro 25	Asp	G 1 y	Phe	A s p	G 1 y 3 0	Leu	G 1 y
Туг	Arg	G 1 y 3 5	G 1 y	Ala	Arg As	р	G 1 u 4 0	Gln	G 1 y	Рһе	Gly	G 1 y 4 5	Ala	Phe	Рго
Ala	Arg 50	Ser	Phe	Ser	Thr Gi		Ser	Asp	Leu	G 1 y	H i s 60	Trp	V a 1	Thr	Thr
Pro 65	Pro	Asp	Ile		Gly Se 70	r	Arg	Asn	Leu	His 75	Trp	G 1 y	Glu	Lys	S c r 8 0
Pro	Pro	Туг	G 1 y	V a 1 :	Pro Ti	ır	Thr	S e r	Thr 90	Pro	Туг	Glu	Gly	Pro 95	Thr
Glu	Glu	Pro	Phe 100	Ser	Ser G	i y	G 1 y	G 1 y 1 0 5	G 1 y	Ser	Vai	Gln	G1 y 110	Gln	Ser
Ser	Glu	G 1 n 1 1 5	Leu	Asn.	Arg Pi	h e	A 1 a 1 2 0	G 1 y	Phe	G 1 y	I 1 c	G 1 y 1 2 5	Leu	Ala	Ser
Leu	Phe 130	Thr	Glu	Asn		e u 3 5	Ala	His	Pro	Суѕ	I 1 e 1 4 0	Val	Leu	Arg	Агд
Gln 145	Сув	Gln	Val		Tyr H 150	is	Ala	Gln	His	T y r 1 5 5	His	Leu	Thr	Pro	Phe 160
Thr	Va1		Asn	I 1 e 1 165	Met T	y r	Ser	Phe	Asn 170	Lys	Thr	Gln	G 1 y	Pro 175	Агд
Ala	Leu		Lys 180	G1 y 1		l y	Ser	Thr 185	Phe	Ilc	Val	Gln	G 1 y 1 9 0	Val	Thr
Leu	Gly	195		Gly	Ile I	l c	Ser 200	Glu	Phe	Thr	Pro	Leu 205	Pro	Arg	Glu
Va 1	Leu 210	His	Lys	Тгр	2	r o 1 5	Lys	Gln	Ile	G 1 y	G 1 u 2 2 0	His	Leu	Leu	Leu
L y s 2 2 5		Leu	Thr		230	a l			Pro	Phe 235	Туг	Ser	Ala	Ser	Leu 240
Ile	Glu		Val	2 4 5		l u	lic	Iie	Arg 250	Asp	Asn	Thr	Gly	I 1 e 2 5 5	Leu
Glu			Lys 260	Glu	•	le		Arg 265	Val	Ile	G1y	Met	G 1 y 2 7 0	V a 1	Pro
His		275			Leu P		280	Leu	Ser	Leu	I 1 c	Phe 285	Рго	Thr	Val
Leu	His 290	Gly	Val	Leu		у г 9 5	Ile	Ilc	Ser	Ser	V a 1 3 0 0	Ile	Gln	Lys	Phe
V a 1 3 0 5		Leu		Leu	3 1 0	•	Lys		Туг	A s n 3 1 5	Ser	His	Leu	Ala	G 1 u 3 2 0
Ser	Thr	Ser	Pro	V a 1 3 2 5	Gln S	e r	Met	Leu	A s p 3 3 0	Ala	Туг	Phe	Pro	G 1 u 3 3 5	Leu
I 1 e	Ala	Asn	Phe	A 1 a	Ala S	e r	Leu	Сув	Ser	Asp	V a 1	Ile	Leu	Туг	Pro

							-cont	inued							
			3 4 0					3 4 5					3 5 0		
Leu	Glu	Thr 355	V a 1	Leu	His	Arg	Leu 360	His	I 1 c	Gln	G 1 y	Thr 365	Агд	Thr	I 1 e
I 1 e	A s p 3 7 0	Asn	Thr	A s p	Leu	G 1 y 3 7 5	Туг	Glu	V a 1	Leu	Pro 380	I 1 e	Asn	Thr	Gln
Туг 385	Glu	G 1 y	M c t	Arg	A s p 3 9 0	Суs	I 1 e	Asn	Thr	I 1 e 3 9 5	Arg	Gln	Glu	Glu	G 1 y 4 0 0
V a 1	Ph c	G 1 y	Phe	T y r 4 0 5	Lys	G l y	Phe	G 1 y	A 1 a 4 1 0	Val	Ile	Ile	Gln	T y r 4 1 5	Thr
Leu	His	A 1 a	A 1 a 4 2 0	V a i	Leu	Gln	Ile	Thr 425	Lys	Ile	Ile	Туг	S e r 4 3 0	Thr	Leu
Leu	Gln														

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 185 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (i i) MOLECULE TYPE: protein
- (v i) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
- (v i i) IMMEDIATE SOURCE:

O

1

...

- (B) CLONE: YS-39(TB2)
- (\mathbf{x} i) SEQUENCE DESCRIPTION: SEQ ID NO:6:

 Glu
 Leu
 Arg
 Arg
 Phe
 Arg
 Phe
 Leu
 His
 Glu
 Lys
 Asn
 Cys
 Met
 Thr

 Asp
 Leu
 Leu
 Ala
 Lys
 Leu
 Glu
 Ala
 Lys
 Thr
 Gly
 Val
 Asn
 Ser
 Phe

 Ile
 Ala
 Leu
 Gly
 Val
 Ile
 Gly
 Val
 Ala
 Phe
 Gly
 Asn
 Phe
 Gly
 Phe
 Gly
 Asn
 Phe
 Gly
 Phe
 Gly
 Tyr
 Pro
 Asn
 Lys
 Ala
 Asn
 Phe
 Gly
 Tyr
 Pro
 Asn
 Lys
 Asn
 Pro
 Asn
 Lys
 Gly
 Tyr
 Pro
 Asn
 Lys
 Gly
 Tyr
 Pro
 Asn
 Lys
 Gly
 Tyr
 Pro
 Asn
 Lys
 Asn
 Ins
 Ins

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2842 amino acids
 - (B) TYPE: amino acid

- (C) STRANDEDNESS: single (D) TOPOLOGY: linear
- (i i) MOLECULE TYPE: protein
- (v i) ORIGINAL SOURCE:
 (A) ORGANISM: Homo sapiens
- (v i i) IMMEDIATE SOURCE: (B) CLONE: APC

- (\star i) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met 1	Ala	Ala	Ala	Ser 5	Туг	Asp	G 1 n	Leu	Leu 10	Lys	G ļ n	V a l	Glu	A 1 a 1 5	Leu
Lys	Met	Glu	A s n 2 0	Ser	Asn	Leu	Arg	G 1 n 2 5	Glu	Leu	G 1 u	Asp	A s n 3 0	Ser	Asn
H i s	Leu	Thr 35	L y s	Leu·	Glu	Thr	G 1 u 4 0	Ala	Ser	Asn	Met	L y s 4 5	G 1 u	V a l	Leu
Lys	G 1 n 5 0	Leu	Gln	Gly		I 1 e 5 5	Glu	A s p	Glu	Ala	Me t 60	Ala	Ser	Ser	G 1 y
G 1 n 65	I 1 e	A s p	Leu	Leu	Glu 70	Arg	Leu	Lys	G 1 u	Leu 75	Asn	Leu	Asp	Ser	S e r 8 0
Asn	Phe	Pro	G1 y	V a 1 8 5	Lys	Leu	Агд	S e r	L y s 90	Met	Ser	Leu	Arg	S e r 9 5	Туг
G 1 y	Ser	Агд	G 1 u 1 0 0	G 1 y	Ser	Val	Ser	S c r 1 0 5	Arg	Ser	G 1 y	Glu	C y s 1 1 0	Ser	Pro
Val	Pro	Me t 115	G 1 y	Ser	Phe	Pro	Arg 120	Arg	Gly	Phe	'V a 1	Asn 125	G 1 y	Ser	Arg
Glu	Ser 130	Thr	Gly	Туг	Leu	G 1 u 135	Glu	Leu	Glu	Lys	G 1 u 1 4 0	Arg	Ser	Leu	Leu
Leu 145	Ala	Азр	Leu	Asp	Lys 150	Glu	Glu	Lys	G 1 u	L y s 1 5 5	A s p	Trp	Туг	Туг	A 1 a 1 6 0
Gln	Leu	Gln	Asn	Leu 165	Thr	Lys	Агд	I 1 e	Asp 170	Ser	Leu	Leu	Thr	G1 u 175	Asn
Phe	Ser	Leu	G1n 180	Thr	Asp	Met	Thr	Arg 185	Агд	G1 n	Leu	Glu	Tyr 190	Glu	Ala
Агд	Gln	I 1 e 1 9 5	Arg	Va1	A 1 a	Met	G 1 u 2 0 0	Glu	Gln	Leu	Gly	Thr 205	Суs	Gln	Asp
Met	G 1 u 2 1 0	Lys	Arg	Ala	Gln	Arg 215	Агд	I 1 c	Ala	Агд	I i c 2 2 0	Gln	Gln	Ile	Glu
L y s 2 2 5	Asp	Ilc	Leu	Arg	I 1 e 2 3 0	Агд	Gin	Leu	Leu	G 1 n 2 3 5	Ser	Gin	Ala	Thr	G 1 u 2 4 0
Ala	Glu	Агд	Ser	S e r 2 4 5	G 1 n	Азп	Lys	His	G 1 u 2 5 0	Thr	Gly	Ser	His	A s p 2 5 5	Ala
Glu	Arg	Gln	Asn 260	Glu	Gly	Gln	Gly	Val 265	G 1 y	G 1 u	Ile	Asn	Met 270	Ala	Thr
Ser	Gly	Asn 275	-	G1n	G 1 y	Ser	Thr 280	Thr	Arg	Met	Asp	His 285	Glu	Thi	Ala
Ser	V a 1 2 9 0	Leu	Ser	Ser	Ser	Ser 295	Thr	His	Ser	Ala	Pro 300	Arg	Arg	Leu	Thr
S e r 3 0 5		Leu	Gly	Thr	L y s 3 1 0	V a 1	Glu	Met	Val	T y r 3 1 5	Ser	Leu	Leu	Ser	Met 320
Leu	Gly	Thr	His	A s p 3 2 5	Lys	Asp	Азр	Met	S e r 3 3 0	Arg	Thr	Leu	Leu	A 1 a 3 3 5	Met
Ser	Ser	Ser	G 1 n 3 4 0		Ser	Суѕ	Ilc	S e r 3 4 5		Arg	Gla	Ser	G 1 y 3 5 0		Leu
Рго	Leu	Leu	I 1 e	Gln	Leu	Leu	His	G 1 y	Asn	Asp	Lys	Asp	Ser	V a i	Leu

			-continued		
	3 5 5		3 6 0	3 6 5	
Leu Gly 370	Asn Ser A	Arg Gly Ser 375	Lys Glu Ala	Arg Ala Arg Ala 380	Ser Ala
Ala Leu 385	His Asn	Ile Ile His 390	Ser Gln Pro	Asp Asp Lys Arg 395	Gly Arg 400
Arg Glu		Val Leu His 405	Leu Leu Glu 410	Gln lie Arg Ala	Tyr Cys 415
Glu Thr	Cys Trp (420	Glu Trp Gln	Glu Ala His 425	Glu Pro Gly Met	.
Asp Lys	Asn Pro M 435	Met Pro Ala	Pro Val Glu 440	His Gln Ile Cys	Pro Ala
Val Cys 450	Val Leu I	Met Lys Leu 455	Ser Phe Asp	Glu Glu His Ar 460	; His Ala
Met Asn 465	Glu Leu (Gly Gly Leu 470	Gin Ala Ile	Ala Glu Leu Leu 475	ı Gln Val 480
Asp Cys		Tyr Gly Leu 485	Thr Asn Asp 490	His Tyr Ser Ile	Thr Leu 495
Arg Arg	Tyr Ala 6 500	Gly Met Ala	Leu Thr Asn 505	Leu Thr Phe Gi	•
Ala Asn	Lys Ala 5	Thr Leu Cys	Ser Met Lys 520	Gly Cys Met Arg 525	g Ala Leu
Val Ala 530	Gln Leu	Lys Ser Glu 535	Ser Glu Asp	Leu Gln Gln Va 540	l lie Ala
Ser Val	Leu Arg.	Asn Leu Ser 550	Trp Arg Ala	Asp Val Asn Ses	r Lys Lys 560
Thr Leu		Val Gly Ser 565	Val Lys Ala 570	Leu Met Glu Cy	s Ala Leu 575
Glu Val	Lys Lys 580	Glu Ser Thr	Leu Lys Ser 585	Val Leu Ser Al.	
	5 9 5		600	Ala Asp Ile Cy 605	
6 1 0		6 1 5		Leu Thr Tyr Ar 620	-
6 2 5		6 3 0	·	Gly Gly Ile Le	6 4 0
Val Ser	Ser Leu	Ile Ala Thr 645	Asn Glu Asp 650	His Arg Gln II	e Leu Arg 655
	660		6 6 5	His Leu Lys Se 67	0
	6 7 5		680	Leu Trp Asn Le 685	
690	•	6 9 5	_	Asp Met Gly Al 700	
705		7 1 0		Lys Met Ile Al 715	720
		7 2 5	7 3 0		7 3 5
	7 4 0		7 4 5	Ser Leu Pro Se 75	0
	7 5 5		760	Leu Asp Ala Gl 765	
Ser Glu 770		Asp Asa Ile 775		Ser Pro Lys Al 780	a Ser His

	19		OU
		-continued	
Arg Ser Lys	Gln Arg His Lys		Tyr Gly Asp Tyr Val Phe
785	790		795 800
Asp Thr Asn	Arg His Asp Asp	Asn Arg Ser A	Asp Asn Phe Asn Thr Gly
	805	810	815
Asn Met Thr	Val Leu Ser Pro	Tyr Leu Asn 7	Thr Thr Val Leu Pro Ser
	820	825	830
Ser Ser Ser	Ser Arg Gly Ser	Leu Asp Ser S	Ser Arg Ser Glu Lys Asp
835		840	845
Arg Ser Leu 850	Glu Arg Glu Arg		Leu Gly Asn Tyr His Pro 860
Ala Thr Glu 865	Asn Pro Gly Th:	•	Arg Gly Leu Gln Ile Ser 875 880
Thr Thr Ala	Ala Gin Ile Ala 885	a Lys Vai Met (Glu Glu Val Ser Ala Ile 895
His Thr Ser	Gln Glu Asp Ar	g Ser Ser Gly S	Ser Thr Thr Glu Leu His
	900	905	910
Cys Val Thr	Asp Glu Arg As	n Ala Leu Arg	Arg Ser Ser Ala Ala His
915		920	925
Thr His Ser 930	Asn Thr Tyr As:	•	Ser Glu Asn Ser Asn Arg 940
Thr Cys Ser	Met Pro Tyr Al		Tyr Lys Arg Ser Ser Asn
945	950		955 960
Asp Ser Leu	Asn Ser Val Se	r Ser Ser Asp	Gly Tyr Gly Lys Arg Gly
	965	970	975
Gln Met Lys	Pro Ser Ile Gl	u Ser Tyr Ser (985	Glu Asp Asp Glu Ser Lys 990
Phe Cys Ser 995		r Pro Ala Asp 1	Leu Ala His Lys Ile His 1005
Ser Ala Asn	His Met Asp As		Glu Leu Asp Thr Pro Ile
1010	10		1020
Asn Tyr Ser	Leu Lys Tyr Se	-	Leu Asn Ser Gly Arg Gln
1025	1030		1035 1040
Ser Pro Ser	Gln Asn Glu Ar	g Trp Ala Arg	Pro Lys His Ile Ile Glu
	1045	1050	1055
Asp Glu Ile	Lys Gln Ser Gl	u Gin Arg Gin	Ser Arg Asn Gln Ser Thr
	1060	1065	1070
Thr Tyr Pro		u Ser Thr Asp	Asp Lys His Leu Lys Phe
107		1080	1085
Gln Pro His 1090	Phe Gly Gln G1		Ser Pro Tyr Arg Ser Arg 1100
Gly Ala Asn	Gly Ser Glu Th		Gly Ser Asn His Gly Ile
1105	1110		1115 1120
Asn Gln Asn	Val Ser Gln Se	r Leu Cys Gln	Glu Asp Asp Tyr Glu Asp
	1125	1130	1135
Asp Lys Pro	Thr Asn Tyr Se	r Glu Arg Tyr	Ser Glu Glu Glu Gln His
	1140	1145	1150
Glu Glu Glu		r Asn Tyr Ser	Ile Lys Tyr Asn Glu Glu
115		1160	1165
Lys Arg His		o Ile Asp Tyr	Ser Leu Lys Tyr Ala Thr
1170		75	1180
Asp Ile Pro	Ser Ser Gln Ly		Ser Phe Ser Lys Ser Ser
1185	1190		1195
Ser Gly Gir	n Ser Ser Lys Th	r Glu His Met	Ser Ser Ser Glu Asn
	1205	1210	1215

						4
-00	n	Ħ	n	17	Α.	п

Thr	Ser	Thr	Pro Ser 1220	Ser Asn		Lys Arg C 1225	31n Asn (Gln Leu His Pro 1230
Ser	S e r	A 1 a		Arg Ser	G1y 6	Gln Pro C		Ala Ala Thr Cys 1245
Lys	V a l 1 2 5 0		Ser Ile	Asn Gln 125		Thr Ile (31n Thr 3 1260	Tyr Cys Val Glu
Asp 1265		Pro	Ile Cys	Phe Ser 1270	Arg		Ser Leu : 1275	Ser Ser Leu Ser 1280
Ser	Ala	Glu	Asp Glu 128		Сув	Asn Gln 7 1290	Thr Thr (Gin Glu Ala Asp 1295
Ser	A 1 a	A s n	Thr Leu 1300	Gln Ile		Glu Ile I 1305	Lys Glu i	Lys Ile Gly Thr 1310
Агд	Ser	A 1 a		Pro Val	Ser 1320	Glu Val I		Val Ser Gln His 1325
Pro	Arg 133		Lys Ser	Ser Arg		Gln Gly S	Ser Ser 1340	Leu Ser Ser Glu
Ser 1345		Arg	His Lys	Ala Val	Glu		Ser Gly . 1355	Ala Lys Ser Pro 1360
Ser	Lys	Ser	Gly Ala 136		Рго	Lys Ser 1 1370	Pro Pro	Glu His Tyr Val 1375
Gln	Glu	Thr	Pro Leu 1380	Met Phe		Arg Cys 1 1385	Thr Ser	Val Ser Ser Leu 1390
Авр	Ser	Phe 139		r Arg Ser	I 1 e 1 4 0 0			Gln Ser Glu Pro 1405
Суз	Ser 141		Met Val	i Ser Gly 141		Ile Ser	Pro Ser 1420	Asp Leu Pro Asp
Ser 142		G 1 y	Gla Thi	Met Pro	Pro		Ser Lys 1435	Thr Pro Pro Pro 1440
Pro	Pro	Gln	Thr A1 a		Lys	Arg Glu 1450	Val Pro	Lys Asn Lys Ala 1455
Pro	Thr	Ala	Glu Ly: 1460	s Arg Glu		Gly Pro 1 1465	Lys Gln	Ala Ala Val Asn 1470
Ala	Ala	V a 1 1 4 7		g Val Glm	Val 1480			Asp Thr Leu Leu 1485
His	Phe 149		Thr Gl	u Ser Thr 149		Asp Gly	Phe Ser 1500	Cys Ser Ser Ser
Leu 150	Ser 5	Ala	Leu Se	r Leu Asp 1510	Glu	Pro Phe	Ile Gln 1515	Lys Asp Val Glu 1520
Leu	Arg	Ile	Met Pro		Gln	Glu Asn 1530		Gly Asn Glu Thr 1535
Glu	Ser	Glu	Gln Pr 1540	o Lys Gla		Asn Glu 1545	Asn Gln	Glu Lys Glu Ala 1550
Glu	Lys	Thr 155		p Ser Gla	1 Lys 1560			Asp Ser Asp Asp 1565
Азр	Asp 157		Giu Il	e Leu Gli 15		Cys Ile	Ile Ser 1580	Ala Met Pro Thr
Lys 158		Ser	Arg Ly	s Ala Ly: 1590	s Lys		Gln Thr 1595	Ala Ser Lys Leu 1600
Рго	Pro	Pro		a Arg Ly 05	s Pro	Ser Gln 1610		Val Tyr Lys Leu 1615
Leu	Pro	Ser	Gln As 1620	n Arg Le	u G1n	Pro Gln 1625	Lys His	Val Ser Phe Thr 1630
Pro	G 1 y	Asp	Asp Me	t Pro Ar	g Val	Туг Сув	Val Glu	Gly Thr Pro Ile

	-continued
3 5	1640

	-continued	
1 6 3 5	1 6 4 0	1 6 4 5
Asn Phe Ser Thr 1650	Ala Thr Ser Leu Ser Asp Leu Th	r Ile Glu Ser Pro i60
Pro Asn Glu Leu	Ala Ala Gly Glu Gly Vai Arg Gl	y Gly Ala Gln Ser
1665	1670 1675	1680
Gly Glu Phe Glu	Lys Arg Asp Thr Ile Pro Thr Gl 1685	u Gly Arg Ser Thr 1695
Asp Glu Ala Gln	Gly Gly Lys Thr Ser Ser Val Th	ar Ile Pro Glu Leu
1700	1705	1710
Asp Asp Asn Lys	Ala Glu Glu Gly Asp Ile Leu Ai	la Glu Cys Ile Asn
1715	1720	1725
Ser Ala Met Pro	Lys Gly Lys Ser His Lys Pro Ph	ne Arg Val Lys Lys
1730	1735	740
Ile Met Asp Gln	Val Gin Gin Ala Ser Ala Ser Se	er Ser Ala Pro Asn
1745	1750 1755	1760
Lys Asn Gln Leu	Asp Gly Lys Lys Lys Lys Pro Th 1765 1770	nr Ser Pro Val Lys 1775
Pro Ile Pro Gln	Asn Thr Glu Tyr Arg Thr Arg Va	al Arg Lys Asn Ala
1780	1785	1790
Asp Ser Lys Asn	Asn Leu Asn Ala Glu Arg Val Ph	ne Ser Asp Asn Lys
1795	1800	1805
Asp Ser Lys Lys	Gln Asn Leu Lys Asn Asn Ser Ly	ys Asp Phe Asn Asp
1810	1815	320
Lys Leu Pro Asn	Asn Glu Asp Arg Val Arg Gly Sc	er Phe Ala Phe Asp
1825	1830 1835	1840
Ser Pro His His	Tyr Thr Pro Ile Glu Gly Thr Pr 1845 1850	ro Tyr Cys Phe Ser 1855
Arg Asn Asp Ser	Leu Ser Ser Leu Asp Phe Asp As	sp Asp Asp Val Asp
1860	0 1865	1870
Leu Ser Arg Glu	Lys Ala Glu Leu Arg Lys Ala Ly	ys Glu Asn Lys Glu
1875	1880	1885
Ser Glu Ala Lys 1890	Val Thr Ser His Thr Glu Leu Th	nr Ser Asn Gln Gln 900
Ser Ala Asn Lys	Thr Gln Ala Ile Ala Lys Gln Pi	ro Ile Asn Arg Gly
1905	1910 1915	1920
Gln Pro Lys Pro	Ile Leu Gln Lys Gin Ser Thr P1 1925 1930	
Lys Asp IIe Pro	Asp Arg Gly Ala Ala Thr Asp Gl	lu Lys Leu Gln Asn
1946	1945	1950
Phe Ala Ile Glu	Asn Thr Pro Val Cys Phe Ser H:	is Asn Ser Ser Leu
1955	1960	1965
Ser Ser Leu Ser 1970	Asp Ile Asp Gln Glu Asn Asn Asn 1975	sn Lys Glu Asn Glu 980
Pro Ile Lys Glu	Thr Glu Pro Pro Asp Ser Gln G	ly Glu Pro Ser Lys
1985	1990 1995	2000
Pro Gln Ala Ser	Gly Tyr Ala Pro Lys Ser Phe H 2005 2010	is Val Glu Asp Thr 2015
Pro Val Cys Phe	Ser Arg Asn Ser Ser Leu Ser Se	er Leu Ser Ile Asp
2020	0 2025	2030
Ser Glu Asp Asp	Leu Leu Gln Glu Cys Ile Ser Se	er Ala Met Pro Lys
2035	2040	2045
Lys Lys Lys Pro	Ser Arg Leu Lys Gly Asp Asn G	lu Lys His Ser Pro
2050	2055	060

Arg 2065		Met	Gly	G 1 y	I 1 e 2 0 7 0		G 1 y	Glu	Asp	Leu 2075		Leu	Asp	Leu	L y s 2 0 8 0
Asp	I 1 e	Gln	Агд	Pro 2085	_	Ser	Glu	His	G 1 y 2 0 9 0		Ser	Pro	A s p	Ser 2095	
Asn	Phe	Азр	Trp 210	-	A 1 a	Ile	G 1 n	G i u 2 1 0 3		Ala	Asn	Ser	I 1 c 2 1 1 c		Ser

Ser Leu His Gln Ala Ala Ala Ala Cys Leu Ser Arg Gin Ala Ser 2120

Ser Asp Ser Asp Ser Ile Leu Ser Leu Lys Ser Gly Ile Ser Leu Gly 2135

Ser Pro Phe His Leu Thr Pro Asp Glu Glu Lys Pro Phe Thr Ser 2 1 5 0 2155

Asn Lys Gly Pro Arg lie Leu Lys Pro Gly Glu Lys Ser Thr Leu Glu 2170

Thr Lys Lys Ile Glu Ser Glu Ser Lys Gly Ile Lys Gly Gly Lys Lys

Val Tyr Lys Ser Leu Ile Thr Gly Lys Val Arg Ser Asn Ser Glu Ile

Ser Gly Gln Met Lys Gln Pro Leu Gln Ala Asn Met Pro Ser Ile Ser 2215

Arg Gly Arg Thr Met Ile His Ile Pro Gly Val Arg Asn Ser Ser 2225 2230 2235

Ser Thr Ser Pro Val Ser Lys Lys Gly Pro Pro Leu Lys Thr Pro Ala 2245 2250 2255

Ser Lys Ser Pro Ser Glu Gly Gln Thr Ala Thr Thr Ser Pro Arg Gly 2265

Ala Lys Pro Ser Val Lys Ser Glu Leu Ser Pro Val Ala Arg Gin Thr 2280 2285

Ser Gln Ile Gly Gly Ser Ser Lys Ala Pro Ser Arg Ser Gly Ser Arg 2295

Asp Ser Thr Pro Ser Arg Pro Ala Gln Pro Leu Ser Arg Pro Ile 2305 2310 2315 232

Gln Ser Pro Gly Arg Asn Ser Ile Ser Pro Gly Arg Asn Gly Ile Ser 2325 2330

Pro Pro Asn Lys Leu Ser Gln Leu Pro Arg Thr Ser Ser Pro Ser Thr 2345

Ala Ser Thr Lys Ser Ser Gly Ser Gly Lys Met Ser Tyr Thr Ser Pro 2360

Gly Arg Gln Met Ser Gln Gln Asn Leu Thr Lys Gln Thr Gly Leu Ser

Lys Asn Ala Ser Ser Ile Pro Arg Ser Glu Ser Ala Ser Lys Gly Leu 2390 2395

Asn Gln Met Asn Asn Gly Asn Gly Ala Asn Lys Lys Val Glu Leu Ser

Arg Met Ser Ser Thr Lys Ser Ser Gly Ser Glu Ser Asp Arg Ser Glu 2420 2430

Arg Pro Val Leu Val Arg Gln Ser Thr Phe Ile Lys Glu Ala Pro Ser

Pro Thr Leu Arg Arg Lys Leu Glu Glu Ser Ala Ser Phe Glu Ser Leu 2 4 5 5

Ser Pro Ser Ser Arg Pro Ala Ser Pro Thr Arg Ser Gln Ala Gln Thr 2465 2470 2475 248

Pro Val Leu Ser Pro Ser Leu Pro Asp Met Ser Leu Ser Thr His Ser 2485 2490

-1 S 11 į ļ. 13 1

-00	nti	-	Δđ

Ser Vai Gln Ala Gly Gly Trp Arg Lys Leu Pro 2500 2505	Pro Asn Leu Ser Pro 2510
Thr Ile Glu Tyr Asn Asp Gly Arg Pro Ala Lys 2515 2520	Arg His Asp Ile Ala 2525
Arg Ser His Ser Glu Ser Pro Ser Arg Leu Pro 2535	Ile Asn Arg Ser Gly 2540
Thr Trp Lys Arg Glu His Ser Lys His Ser Ser 2545 2550 255	
Ser Thr Trp Arg Arg Thr Gly Ser Ser Ser Ser 2565	2 5 7 5
Ser Glu Ser Ser Glu Lys Ala Lys Ser Glu Asp 2580 2585	2 5 9 0
Ser Ile Ser Gly Thr Lys Gln Ser Lys Glu Asn 2595 Gly Thr Trp Arg Lys Ile Lys Glu Asn Glu Phe	2605
2610 2615 Thr Ser Gln Thr Val Ser Ser Gly Ala Thr Asn	2620
2625 2630 2630 263 Thr Leu lle Tyr Gln Met Ala Pro Ala Val Ser	3 5 2 6 4 0
2645 2650 Trp Val Arg Ile Glu Asp Cys Pro Ile Asn Asn	2 6 5 5
2660 2665 Ser Pro Thr Gly Asn Thr Pro Pro Val Ile Asp	2670
2675 2680 Ala Asn Pro Asn Ile Lys Asp Ser Lys Asp Asn	
2690 2695 Val Gly Asn Gly Ser Val Pro Met Arg Thr Val 2705 2710 271	
2705 2710 271 Leu Asn Ser Phe Ile Gln Val Asp Ala Pro Asp 2725 2730	
Ile Lys Pro Gly Gln Asn Asn Pro Val Pro Val 2740 2745	
Ser Ser Ile Val Glu Arg Thr Pro Phe Ser Ser 2755	r Ser Ser Ser Lys 2765
His Ser Ser Pro Ser Gly Thr Val Ala Ala Arg 2770 2775	g Val Thr Pro Phe Asn 2780
Tyr Asa Pro Ser Pro Arg Lys Ser Ser Ala Asg 2785 2790 279	
Pro Ser Gln Ile Pro Thr Pro Val Asn Asn Ass 2805 2810	n Thr Lys Lys Arg Asp 2815
Ser Lys Thr Asp Ser Thr Glu Ser Ser Gly The	r Gln Ser Pro Lys Arg 2830
His Ser Gly Ser Tyr Leu Val Thr Ser Val 2835 2840	

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- $(\ i\ i\)$ MOLECULE TYPE: peptide
- (v i i) IMMEDIATE SOURCE:
 - (B) CLONE: ral2(yeast)

```
-continued
```

```
( x i ) SEQUENCE DESCRIPTION: SEQ ID NO:8:
       Leu Thr Gly Ala Lys Gly Leu Gln Leu Arg Ala Leu Arg Arg Ile Ala
1 10 15
       Arg Ile Glu Gln Gly Gly Thr Ala Ile Ser Pro Thr Ser Pro Leu
( \,2\, ) INFORMATION FOR SEQ ID NO:9:
       ( i ) SEQUENCE CHARACTERISTICS:
               ( A ) LENGTH: 29 amino acids
                (B) TYPE: amino acid
                ( C ) STRANDEDNESS: single
               ( D ) TOPOLOGY: linear
      ( i i ) MOLECULE TYPE: peptide
      ( v i ) ORIGINAL SOURCE:
               ( A ) ORGANISM: Homo sapiens
     ( v i i ) IMMEDIATE SOURCE:
               (B) CLONE: m3(mAChR)
      ( \mathbf{x} i ) SEQUENCE DESCRIPTION: SEQ ID NO:9:
       Leu Tyr Trp Arg Ile Tyr Lys Giu Thr Glu Lys Arg Thr Lys Glu Leu
       Ala Gly Leu Gln Ala Ser Gly Thr Glu Ala Glu Thr Glu 20\,
( 2 ) INFORMATION FOR SEQ ID NO:10:
        ( i ) SEQUENCE CHARACTERISTICS:
                (A) LENGTH: 29 amino acids
                (B) TYPE: amino acid
                (C) STRANDEDNESS: single
                ( D ) TOPOLOGY: linear
      ( i i ) MOLECULE TYPE: peptide
      ( v i ) ORIGINAL SOURCE:
                ( A ) ORGANISM: Homo sapiens
     ( v i i ) IMMEDIATE SOURCE:
                (B) CLONE: MCC
       ( \mathbf{x} i ) SEQUENCE DESCRIPTION: SEQ ID NO:10:
        Leu Tyr Pro Asn Leu Ala Glu Glu Arg Ser Arg Trp Glu Lys Glu Leu
        Ala Gly Leu Arg Glu Glu Asa Glu Ser Leu Thr Ala Met
( 2 ) INFORMATION FOR SEQ ID NO:11:
        ( i ) SEQUENCE CHARACTERISTICS:
                ( A ) LENGIH: 40 base pairs
                (B) TYPE: nucleic acid
                ( C ) STRANDEDNESS: single
                ( D ) TOPOLOGY: linear
       ( i i ) MOLECULE TYPE: cDNA
       ( v i ) ORIGINAL SOURCE:
                ( A ) ORGANISM: Homo sapiens
       ( \mathbf{x}\ i ) SEQUENCE DESCRIPTION: SEQ ID NO:11:
GTATCAAGAC TGTGACTTTT AATTGTAGTT TATCCATTTT
```

(2) INFORMATION FOR SEQ ID NO:12:

聪

H

```
-continued
         ( 1 ) SEQUENCE CHARACTERISTICS:
                 ( A ) LENGTH: 40 base pairs
                  ( B ) TYPE: nucleic acid
                 ( C ) STRANDEDNESS: single
                  ( D ) TOPOLOGY: linear
       ( i i ) MOLECULE TYPE: cDNA
       ( v 1 ) ORIGINAL SOURCE:
                 ( A ) ORGANISM: Homo sapiens
       ( x i ) SEQUENCE DESCRIPTION: SEQ ID NO:12:
TTTAGAATTT CATGTTAATA TATTGTGTTC TTTTTAACAG
                                                                                                                           4 0
( 2 ) INFORMATION FOR SEQ ID NO:13:
         ( i ) SEQUENCE CHARACTERISTICS:
                 ( A ) LENGTH: 40 base pairs
                 (B) TYPE: nucleic acid
                  ( C ) STRANDEDNESS: single
                  ( D ) TOPOLOGY: linear
       ( i i ) MOLECULE TYPE: cDNA
       ( v i ) ORIGINAL SOURCE:
                 ( A ) ORGANISM: Homo sapiens
       ( x i ) SEQUENCE DESCRIPTION: SEQ ID NO:13:
GTAGATTTTA AAAAGGTGTT TTAAAATAAT TTTTTAAGCT
( 2 ) INFORMATION FOR SEQ ID NO:14:
         ( i ) SEQUENCE CHARACTERISTICS:
                  ( A ) LENGTH: 40 base pairs
                  ( B ) TYPE: nucleic acid
                  ( C ) STRANDEDNESS: single
                  ( D ) TOPOLOGY: linear
       ( i i ) MOLECULE TYPE: cDNA
       ( v i ) ORIGINAL SOURCE:
                  ( A ) ORGANISM: Homo sapiens
       ( \mathbf{x} i ) SEQUENCE DESCRIPTION: SEQ ID NO:14:
AAGCAATTGT TGTATAAAAA CTTGTTTCTA TTTTATTTAG
                                                                                                                           4 0
( 2 ) INFORMATION FOR SEQ ID NO:15:
         ( i ) SEQUENCE CHARACTERISTICS:
                  ( A ) LENGTH: 40 base pairs
                  ( B ) TYPE: nucleic acid
                  ( C ) STRANDEDNESS: single
                  ( D ) TOPOLOGY: linear
       ( i i ) MOLECULE TYPE: cDNA
       ( v i ) ORIGINAL SOURCE:
                  ( A ) ORGANISM: Homo sapiens
       ( x i ) SEQUENCE DESCRIPTION: SEQ ID NO:15:
GTAACTTTC TTCATATAGT AAACATTGCC TTGTGTACTC
                                                                                                                           4 0
( 2 ) INFORMATION FOR SEQ ID NO:16:
         ( i ) SEQUENCE CHARACTERISTICS:
                  ( A ) LENGTH: 40 base pairs
                  (B) TYPE: nucleic acid
                  ( C ) STRANDEDNESS: single
                  (D) TOPOLOGY: linear
       ( i i ) MOLECULE TYPE: cDNA
```

```
-continued
```

	-continuou	
(v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens		
($\mathbf{x} \;\; \mathbf{i} \;\;)$ SEQUENCE DESCRIPTION: SEQ ID NO:16:		
NNNNNNNNN NNNGTCCCTT TTTTTAAAAA	AAAAAATAG	4 0
(2) INFORMATION FOR SEQ ID NO:17:		
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 40 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear		
(i i) MOLECULE TYPE: cDNA		
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens		
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:17:		
GTAAGTAACT TGGCAGTACA ACTTATTTGA	AACTTTAATA	4 0
(2) INFORMATION FOR SEQ ID NO:18:		
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 40 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear		
(i i) MOLECULE TYPE: cDNA		
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens		
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:18:		
ATACAAGATA TTGATACTTT TTTATTATTT	GTGGTTTTAG	4 0
(2) INFORMATION FOR SEQ ID NO:19:		
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 40 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear		
(i i) MOLECULE TYPE: cDNA		
(v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens		
(\mathbf{x} i) SEQUENCE DESCRIPTION: SEQ ID NO:19:		
GTAAGTTACT TGTTTCTAAG TGATAAAACA	G Y GAAGAGCT	4 0
(2) INFORMATION FOR SEQ ID NO.20:		
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 40 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear		
(i i) MOLECULE TYPE: cDNA		
(v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sepiens		
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:20:		
AATAAAAACA TAACTAATTA GGTTTCTTGT	TTTATTTAG	4.0

-continued

(2) INFORMATION FOR SEQ ID NO.21:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 40 base pairs	
(B) TYPE: nucleic acid (C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:21:	
GTTAGTAAAT TSCCTTTTTT GTTTGTGGGT ATAAAAATAG	4 0
(2) INFORMATION FOR SEQ ID NO:22:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 40 base pairs	
(B) TYPE: muleic acid (C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:22:	
ACCATTTTTG CATGTACTGA TGTTAACTCC ATCTTAACAG	4 0
(2) INFORMATION FOR SEQ ID NO:23:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 40 base pairs	
(B) TYPE: mucleic acid	
(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: eDNA	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(\mathbf{x} i) SEQUENCE DESCRIPTION: SEQ ID NO:23:	
GTAAATAAAT TATTTTATCA TATTTTTTAA AATTATTTAA	4 0
(2) INFORMATION FOR SEQ ID NO:24:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 64 base pairs	
(B) TYPE: mucleic acid	
(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO.24: CATGATGTTA TCTGTATTTA CCTATAGTCT AAATTATACC ATCTATAATG TGCTTAATTT	£ 0
TTAG	60
	04
(2) INFORMATION FOR SEQ ID NO.25:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 52 base pairs (B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	

	conunaca		
(D) TOPOLOGY: linear			
(i i) MOLECULE TYPE: cDNA			
(vi)ORIGINAL SOURCE: (A)ORGANISM: Homo sapiens			
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:25:			
GTAACAGAAG ATTACAAACC CTGGTCACTA	ATGCCATGAC	TACTTTGCTA AG	5 2
(2) INFORMATION FOR SEQ ID NO:26:			
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 46 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear			
(i i) MOLECULE TYPE: cDNA			
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens			
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:26:			
GGATATTAAA GTCGTAATTT TGTTTCTAAA	CTCATTTGGC	CCACAG	4 6
(2) INFORMATION FOR SEQ ID NO:27:			
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 40 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear			
(i i) MOLECULE TYPE: cDNA			
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens			
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:27:			
GTATGTTCTC TATAGTGTAC ATCGTAGTGC	ATGTTTCAAA		4 0
(2) INFORMATION FOR SEQ ID NO:28:			
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 56 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear			
(i i) MOLECULE TYPE: cDNA			
(v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens			
(\mathbf{x} i) SEQUENCE DESCRIPTION: SEQ ID NO:28:			
CATCATTGCT CTTCAAATAA CAAAGCATTA	TGGTTTATGT	TGATTTTATT TTTCAG	5 6
(2) INFORMATION FOR SEQ ID NO:29:			
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 43 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear			
(i i) MOLECULE TYPE: cDNA			
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens			
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:29:			

```
-continued
GTAAGACAAA AATGTTTTTT AATGACATAG ACAATTACTG GTG
                                                                                                                   4 3
(2) INFORMATION FOR SEQ ID NO:30:
        ( 1 ) SEQUENCE CHARACTERISTICS:
                ( A ) LENGTH: 40 base pairs
                ( B ) TYPE: nucleic acid
                ( C ) STRANDEDNESS: single
                ( D ) TOPOLOGY: linear
      ( i i ) MOLECULE TYPE: cDNA
      ( v i ) ORIGINAL SOURCE:
                ( A ) ORGANISM: Homo sapiens
      ( x 1 ) SEQUENCE DESCRIPTION: SEQ ID NO:30:
TTAGATGATT GTCTTTTTCC TCTTGCCCTT TTTAAATTAG
                                                                                                                   40
(2) INFORMATION FOR SEQ ID NO:31:
        (\ i\ ) SEQUENCE CHARACTERISTICS:
                 ( A ) LENGTH: 44 base pairs
                 (B) TYPE: nucleic acid
                 ( C ) STRANDEDNESS: single
                 ( D ) TOPOLOGY: linear
       ( i i ) MOLECULE TYPE: cDNA
       ( v i ) ORIGINAL SOURCE:
                 ( A ) ORGANISM: Homo sapiens
       ( \mathbf{x} i ) SEQUENCE DESCRIPTION: SEQ ID NO:31:
GTATGTTTT ATAACATGTA TTTCTTAAGA TAGCTCAGGT ATGA
( 2 ) INFORMATION FOR SEQ ID NO:32:
        ( i ) SEQUENCE CHARACTERISTICS:
                 ( A ) LENGTH: 54 base pairs
                 ( B ) TYPE: nucleic acid
                 ( C ) STRANDEDNESS: single
                 ( D ) TOPOLOGY: linear
       ( i i ) MOLECULE TYPE: cDNA
       ( v i ) ORIGINAL SOURCE:
                ( A ) ORGANISM: Homo sapiens
       ( x i ) SEQUENCE DESCRIPTION: SEQ ID NO:32:
GCTTGGCTTC AAGTTGNCTT TTTAATGATC CTCTATTCTG TATTTAATTT ACAG
                                                                                                                   5 4
( 2 ) INFORMATION FOR SEQ ID NO:33:
         ( i ) SEQUENCE CHARACTERISTICS:
                 ( A ) LENGTH: 65 base pairs
                 ( B ) TYPE: nucleic acid
                 (C) STRANDEDNESS: single
                 ( D ) TOPOLOGY: linear
       ( i i ) MOLECULE TYPE: cDNA
       ( v i ) ORIGINAL SOURCE:
                 ( A ) ORGANISM: Homo sapiens
       ( x i ) SEQUENCE DESCRIPTION: SEQ ID NO:33:
GTACTATTTA GAATTTCACC TGTTTTTCTT TTTTCTCTTT TTCTTTTGAGG CAGGGTCTCA
                                                                                                                    60
CTCTG
                                                                                                                    65
( 2 ) INFORMATION FOR SEQ ID NO:34:
```

(i i) MOLECULE TYPE: cDNA

```
-continued
```

	Condition
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 52 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(i 1) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:34:	
GCAACTAGTA TGATTTTATG TATAAATTAA	TCTAAAATTG ATTAATTTCC AG 52
(2) INFORMATION FOR SEQ ID NO:35:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 42 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
($\mathbf{x} \ \mathbf{i} \)$ SEQUENCE DESCRIPTION: SEQ ID NO:35:	
GTACCTTTGA AAACATTTAG TACTATAATA	TGAATTTCAT GT 42
(2) INFORMATION FOR SEQ ID NO:36:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 40 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:36:	
CCAACTCNAA TTAGATGACC CATATTCAGA	AACTTACTAG 40
(2) INFORMATION FOR SEQ ID NO:37:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 54 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:37:	
GTATATATAG AGTTTTATAT TACTTTTAAA	GTACAGAATT CATACTCTCA AAAA 54
(2) INFORMATION FOR SEQ ID NO:38:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 41 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	

103 104 -continued (v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens ($\mathbf{x}\ i$) SEQUENCE DESCRIPTION: SEQ ID NO:38: ATTGTGACCT TAATTTTGTG ATCTCTTGAT TTTTATTTCA G (2) INFORMATION FOR SEQ ID NO:39: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 18 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (i i) MOLECULE TYPE: cDNA (v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens (\mathbf{x} i) SEQUENCE DESCRIPTION: SEQ ID NO:39: TCCCCGCCTG CCGCTCTC 18 (2) INFORMATION FOR SEQ ID NO:40: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 18 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (i i) MOLECULE TYPE: cDNA (v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens (x i) SEQUENCE DESCRIPTION: SEQ ID NO:40: GCAGCGGCGG CTCCCGTG (2) INFORMATION FOR SEQ ID NO:41: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (i i) MOLECULE TYPE: cDNA (v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens (x i) SEQUENCE DESCRIPTION: SEQ ID NO:41: GTGAACGGCT CTCATGCTGC 20 (2) INFORMATION FOR SEQ ID NO:42: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 19 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single

- (D) TOPOLOGY: linear
- (i i) MOLECULE TYPE: cDNA
- (v i) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
- ($\mathbf{x} \;\; \mathbf{i} \;\;)$ SEQUENCE DESCRIPTION: SEQ ID NO:42:

ACGTGCGGGG AGGAATGGA

(2) INFORMATION FOR SEQ ID NO:43:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 24 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
($\mathbf{x} \ \mathbf{i} \)$ SEQUENCE DESCRIPTION: SEQ ID NO:43:	
ATGATATOTT ACCAAATGAT ATAC	2 4
(2) INFORMATION FOR SEQ ID NO:44:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 23 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(\mathbf{x} i) SEQUENCE DESCRIPTION: SEQ ID NO:44:	
TTATTCCTAC TTCTTCTATA CAG	2 3
	23
(2) INFORMATION FOR SEQ ID NO:45:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 21 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(v i) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sepiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:45:	
TACCCATGCT GGCTCTTTTT C	2 1
(2) INFORMATION FOR SEQ ID NO:46:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 20 base pairs	
(B) TYPE: mucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:46:	
TGGGGCCATC TTGTTCCTGA	2 0
(0) 117001 (4770) 1700 070 17 17 17	
(2) INFORMATION FOR SEQ ID NO:47:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 22 base pairs	
(B) TYPE: mucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	

(i i) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:47:	
ACATTAGGCA CAAAGCTTGC AA	2 2
(2) INFORMATION FOR SEQ ID NO:48:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:48:	
ATCAAGCTCC AGTAAGAAGG TA	2 2
(2) INFORMATION FOR SEQ ID NO:49:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 19 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:49:	
TGCGGCTCCT GGGTTGTTG	1 9
(2) INFORMATION FOR SEQ ID NO:50:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: mucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
($\mathbf{x} \; \mathbf{i} \;$) SEQUENCE DESCRIPTION: SEQ ID NO:50:	
GCCCCTTCCT TTCTGAGGAC	2 0
(2) INFORMATION FOR SEQ ID NO:51:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: modeic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:51:	

(2) INFORMATION FOR SEQ ID NO:52:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs	
(B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
($\mathbf{x} \ \mathbf{i} \)$ SEQUENCE DESCRIPTION: SEQ ID NO:52:	
ATGACACCCC CCATTCCCTC	2 0
(2) INFORMATION FOR SEQ ID NO:53:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 24 base pairs (B) TYPE: mucleic acid	
(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(\mathbf{x} i) SEQUENCE DESCRIPTION: SEQ ID NO:53:	
CCACTTAAAG CACATATATT TAGT	2 4
(2) INFORMATION FOR SEQ ID NO:54:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pairs	
(B) TYPE: mucleic acid	
(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(v i) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:54:	
GTATGGAAAA TAGTGAAGAA CC	2 2
(2) INFORMATION FOR SEQ ID NO:55:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs	
(B) TYPE: nucleic acid	
(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	•
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:55:	
TTCTTAAGTC CTGTTTTTCT TTTG	2 4
(2) INFORMATION FOR SEQ ID NO:56:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGIH: 23 base pairs	
(B) TYPE: nucleic acid (C) STRANDEDNESS: single	

(D) TOPOLOGY: linear

(i i) MOLECULE TYPE: cDNA

(D) TOPOLOGY: linear

(A) ORGANISM: Homo sapiens (x i) SEQUENCE DESCRIPTION: SEQ ID NO:60:

(i i) MOLECULE TYPE: cDNA (v i) ORIGINAL SOURCE:

(v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens (x i) SEQUENCE DESCRIPTION: SEQ ID NO:56: TTTAGAACCT TTTTTGTGTT GTG 23 (2) INFORMATION FOR SEQ ID NO:57: 2 4 22 20 (A) LENGTH: 27 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 23 base pairs

113 114 -continued TAAAAATGGA TAAACTACAA TTAAAAG 27 (2) INFORMATION FOR SEQ ID NO:61: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: mucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (i i) MOLECULE TYPE: cDNA (v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens (x i) SEQUENCE DESCRIPTION: SEQ ID NO:61: AAATACAGAA TCATGTCTTG AAGT 2 4 (2) INFORMATION FOR SEQ ID NO:62: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 23 base pairs (B) TYPE: micleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (i i) MOLECULE TYPE: cDNA (v $\,i\,$) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens (\mathbf{x} i) SEQUENCE DESCRIPTION: SEQ ID NO:62: ACACCTAAAG ATGACAATTT GAG 2 3 (2) INFORMATION FOR SEQ ID NO:63: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (i i) MOLECULE TYPE: cDNA (v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens (x i) SEQUENCE DESCRIPTION: SEQ ID NO:63: TAACTTAGAT AGCAGTAATT TCCC 24 (2) INFORMATION FOR SEQ ID NO:64: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 23 base pairs (B) TYPE: mucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (i i) MOLECULE TYPE: cDNA (v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens (x i) SEQUENCE DESCRIPTION: SEQ ID NO:64: ACAATAAACT GGAGTACACA AGG 23

(A) ORGANISM: Homo sapiens

-continued (B) TYPE: nucleac acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (i i) MOLECULE TYPE: cDNA (v 1) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens (x 1) SEQUENCE DESCRIPTION: SEQ ID NO:65: ATAGGTCATT GCTTCTTGCT GAT 23 (2) INFORMATION FOR SEQ ID NO:66: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (i i) MOLECULE TYPE: cDNA (v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens (x i) SEQUENCE DESCRIPTION: SEQ ID NO:66: TGAATTTTAA TGGATTACCT AGGT 2 4 (2) INFORMATION FOR SEQ ID NO:67: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 25 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (i i) MOLECULE TYPE: cDNA (v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens (x i) SEQUENCE DESCRIPTION: SEQ ID NO:67: CTTTTTTGC TTTTACTGAT TAACG 2 5 (2) INFORMATION FOR SEQ ID NO:68: $(\ i\)$ SEQUENCE CHARACTERISTICS: (A) LENGTH: 27 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (i i) MOLECULE TYPE: cDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens (x i) SEQUENCE DESCRIPTION: SEQ ID NO:68: TGTAATTCAT TTTATTCCTA ATAGCTC 27 (2) INFORMATION FOR SEQ ID NO:69: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (i i) MOLECULE TYPE: cDNA (v i) ORIGINAL SOURCE:

(2) INFORMATION FOR SEQ ID NO:74:

118 -continued (x i) SEQUENCE DESCRIPTION: SEQ ID NO:69: GGTAGCCATA GTATGATTAT TTCT 2 4 (2) INFORMATION FOR SEQ ID NO:70: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (i i) MOLECULE TYPE: cDNA (v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens (x i) SEQUENCE DESCRIPTION: SEQ ID NO:70: CTACCTATTT TTATACCCAC AAAC (2) INFORMATION FOR SEQ ID NO:71: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 23 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (i i) MOLECULE TYPE: cDNA ($v\ i$) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens (x i) SEQUENCE DESCRIPTION: SEQ ID NO:71: AAGAAAGCCT ACACCATTTT TGC 23 (2) INFORMATION FOR SEQ ID NO:72: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 23 base pairs (B) TYPE: mucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (i i) MOLECULE TYPE: cDNA (v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens (x i) SEQUENCE DESCRIPTION: SEQ ID NO:72: GATCATTCTT AGAACCATCT TGC 23 (2) INFORMATION FOR SEQ ID NO:73: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (i i) MOLECULE TYPE: cDNA (v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens (x i) SEQUENCE DESCRIPTION: SEQ ID NO:73: ACCTATAGTC TAAATTATAC CATC

119	120	
	-continued	
() CECUTATOE CITATO A CHEMICAL		
(1) SEQUENCE CHARACTERISTICS:		
(A) LENGTH: 20 base pairs		
(B) TYPE: nucleic acid (C) STRANDEDNESS: single		
(D) TOPOLOGY: linear		
(1 1) MOLECULE TYPE: cDNA		
(= ') ODYGDIAL COLDGE		
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sspiens		
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:74:		
GTCATGGCAT TAGTGACCAG		2 0
(2) INFORMATION FOR SEQ ID NO:75:		
(i) SEQUENCE CHARACTERISTICS:		
(A) LENGTH: 24 base pairs		
(B) TYPE: mucleic acid		
(C) STRANDEDNESS: single		
(D) TOPOLOGY: linear		
(i i) MOLECULE TYPE: cDNA		
(v i) ORIGINAL SOURCE:		
(A) ORGANISM: Homo sapiens		
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:75:		
AGTCGTAATT TTGTTTCTAA ACTC		2 4
(2) INFORMATION FOR SEQ ID NO:76:		
(i) SEQUENCE CHARACTERISTICS:		
(A) LENGTH: 21 base pairs		
(B) TYPE: nucleic acid		
(C) STRANDEDNESS: single		
(D) TOPOLOGY: linear		
(i i) MOLECULE TYPE: cDNA		
/ w : \ODIGINAL SOUDCE.		
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens		
(\mathbf{x} i) SEQUENCE DESCRIPTION: SEQ ID NO:76:		
TGAAGGACTC GGATTTCACG C		2 1
(2) INFORMATION FOR SEQ ID NO:77:		
(i) SEQUENCE CHARACTERISTICS:		
(A) LENGTH: 23 base pairs		
(B) TYPE: nucleic acid		
(C) STRANDEDNESS: single		
(D) TOPOLOGY: linear		
(i i) MOLECULE TYPE: cDNA		
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens		
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:77:		
TCATTCACTC ACAGCCTGAT GAC		2 3
(2) INFORMATION FOR SEQ ID NO:78:		
/ : \ CECTIBACE CITADA CIPEDIOTICO.		

- - (A) LENGTH: 22 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
- (i i) MOLECULE TYPE: cDNA

-continued (v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens (x i) SEQUENCE DESCRIPTION: SEQ ID NO:78: GCTTTGAAAC ATGCACTACG AT (2) INFORMATION FOR SEQ ID NO:79: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (i i) MOLECULE TYPE: cDNA (v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens (x i) SEQUENCE DESCRIPTION: SEQ ID NO:79: AAACATCATT GCTCTTCAAA TAAC 24 (2) INFORMATION FOR SEQ ID NO:80: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: mucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (i i) MOLECULE TYPE: cDNA (v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens (\mathbf{x} i) SEQUENCE DESCRIPTION: SEQ ID NO:80: TACCATGATT TAAAAATCCA CCAG 2 4 (2) INFORMATION FOR SEQ ID NO:81: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 23 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (i i) MOLECULE TYPE: cDNA (v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens (* i) SEQUENCE DESCRIPTION: SEQ ID NO:81: GATGATTGTC TTTTTCCTCT TGC 2 3 (2) INFORMATION FOR SEQ ID NO:82: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: mcleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear (i i) MOLECULE TYPE: cDNA (v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens (x i) SEQUENCE DESCRIPTION: SEQ ID NO:82: CTGAGCTATC TTAAGAAATA CATG 2 4

		•
-con	tını	1eri

(2) INFORMATION FOR SEQ ID NO:83:		
(i) SEQUENCE CHARACTERISTICS:		
(A) LENGTH: 25 base pairs		
(B) TYPE: nucleic acid		
(C) STRANDEDNESS: single		
(D) TOPOLOGY: linear		
(i i) MOLECULE TYPE: cDNA		
(11) Moddoodd 1115, what		
(v i) ORIGINAL SOURCE:		
(A) ORGANISM: Homo sapiens	,	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:83:	•	
(1.1)02402.02.0200		
TTTTAAATGA TCCTCTATTC TGTAT	2	. 5
(a) throps (attrost for one to stood		
(2) INFORMATION FOR SEQ ID NO:84:		
(i) SEQUENCE CHARACTERISTICS:		
(A) LENGTH: 24 base pairs		
(B) TYPE: nucleic acid		
(C) STRANDEDNESS: single		
(D) TOPOLOGY: linear		
(i i) MOLECULE TYPE: cDNA		
(• •) • • • • • • • • • • • • • • • •		
(vi) ORIGINAL SOURCE:		
(A) ORGANISM: Homo sapiens		
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:84:		
ACAGAGTCAG ACCCTGCCTC AAAG	2	4
(2) INFORMATION FOR SEQ ID NO:85:		
() SECTION OF A PARTICULAR		
(i) SEQUENCE CHARACTERISTICS:		
(A) LENGTH: 23 base pairs (B) TYPE: nucleic acid		
(C) STRANDEDNESS: single		
(D) TOPOLOGY: linear		
(i i) MOLECULE TYPE: cDNA		
(v i) ORIGINAL SOURCE:		
(A) ORGANISM: Homo sapiens		
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:85:		
TTTCTATTCT TACTGCTAGC ATT		2 3
(2) INFORMATION FOR SEQ ID NO:86:		
(b) in old all of the sage is notice.		
(i) SEQUENCE CHARACTERISTICS:		
(A) LENGTH: 22 base pairs		
(B) TYPE: nucleic acid		
(C) STRANDEDNESS: single		
(D) TOPOLOGY: linear		
(i i) MOLECULE TYPE: cDNA		
(vi) ORIGINAL SOURCE:		
(A) ORGANISM: Homo sapiens		
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:86:		
ATAGAGAGGT		
ATACACAGGT AAGAAATTAG GA		2 2
_		
(2) INFORMATION FOR SEQ ID NO:87:		
(i) SEQUENCE CHARACTERISTICS:		
(A) LENGTH: 22 base pairs		
(B) TYPE: nucleic acid		
(C) STRANDEDNESS: single		
(D) TOPOLOGY: linear		

	-continued
(i i) MOLECULE TYPE: cDNA	
(v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	r
(x i) SEQUENCE DESCRIPTION: SEQ ID !	VO:87:

TAGATGACCC ATATTCTGTT TC

(2) INFORMATION FOR SEQ ID NO:88:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: mucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (i i) MOLECULE TYPE: cDNA
- (v i) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
- (\mathbf{x} i) SEQUENCE DESCRIPTION: SEQ ID NO:88:

CAATTAGGTC TTTTTGAGAG TA

2 2

22

- (2) INFORMATION FOR SEQ ID NO:89:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 22 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (i i) MOLECULE TYPE: cDNA
 - (v i) ORIGINAL SOURCE:

#

u

- (A) ORGANISM: Homo sapiens
- (x i) SEQUENCE DESCRIPTION: SEQ ID NO:89:

GTTACTGCAT ACACATTGTG AC

22

- (2) INFORMATION FOR SEQ ID NO:90:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23 base pairs
 - (B) TYPE: nucleic acid (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (i i) MOLECULE TYPE: cDNA
 - ($v\ i$) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
 - (x i) SEQUENCE DESCRIPTION: SEQ ID NO:90:

GCTTTTTGTT TCCTAACATG AAG

2 3

- (2) INFORMATION FOR SEQ ID NO:91:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs (B) TYPE: modeic acid

 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (i i) MOLECULE TYPE: cDNA
 - (v i) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
 - (x i) SEQUENCE DESCRIPTION: SEQ ID NO:91:

TCTCCCACAG GTAATACTCC C

(2) INFORMATION FOR SEQ ID NO:92:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid	
(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(x 1) SEQUENCE DESCRIPTION: SEQ ID NO:92:	
GCTAGAACTG AATGGGGTAC G	2 1
(2) INFORMATION FOR SEQ ID NO:93:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 22 base pairs	
(B) TYPE: nucleic acid (C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(\mathbf{x} i) SEQUENCE DESCRIPTION: SEQ ID NO:93:	
CAGGACAAAA TAATCCTGTC CC	2 2
(2) INFORMATION FOR SEQ ID NO:94:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 24 base pairs	
(B) TYPE: nucleic acid (C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(v i) ORIGINAL SOURCE:	
(A) ORGANISM: Homo sapiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:94:	
ATTTTCTTAG TTTCATTCTT CCTC	2 4
(2) INFORMATION FOR SEQ ID NO: 95:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 25 base pairs (B) TYPE: nucleic acid	
(C) STRANDEDNESS: single	
(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sepiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:95:	
AGAAGGATCC CTTGTGCAGT GTGGA	2 :
(2) INFORMATION FOR SEQ ID NO: 96:	
(i) SEQUENCE CHARACTERISTICS:	
(A) LENGTH: 24 base pairs	
(B) TYPE: nucleic acid (C) STRANDEDNESS: single	

($\mathbf{x}\ \mathbf{i}\)$ SEQUENCE DESCRIPTION: SEQ ID NO:100:

(D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(\mathbf{x} i) SEQUENCE DESCRIPTION: SEQ ID NO:96:	
GACAGGATCC TGAAGCTGAG TTTG	2 4
(2) INFORMATION FOR SEQ ID NO: 97:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 18 base pairs (B) TYPE: mucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:97:	
TCAGAAAGTG CTGAAGAG	18
(2) INFORMATION FOR SEQ ID NO: 98:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 19 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(v i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:98:	
GGAATAATTA GGTCTCCAA	19
(2) INFORMATION FOR SEQ ID NO: 99:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
($\mathbf{x} \ \mathbf{i} \)$ SEQUENCE DESCRIPTION: SEQ ID NO.99:	
GCAAATCCTA AGAGAGAACA A	2 1
(2) INFORMATION FOR SBQ ID NO: 100:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 19 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(v i) ORIGINAL SOURCE:	

-continued

GATGGCAAGC TTGAGCCAG	1 9
(2) INFORMATION FOR SEQ ID NO: 101:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 18 base pairs (B) TYPE: mucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:101:	
GTTCCAGCAG TGTCACAG	1 8
(2) INFORMATION FOR SEQ ID NO: 102:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 18 base pairs (B) TYPE: mucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(i i) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(x i) SEQUENCE DESCRIPTION: SEQ ID NO:102:	
GGGAGATTTC GCTCCTGA	1 8

We claim:

1. A preparation of antibodies which specifically binds to a human APC (adenomatous polyposis coli) protein having an amino acid sequence as shown in SEQ ID NO:1, 2, or 7, and does not specifically bind to other human proteins.

2. A preparation of antibodies which specifically binds to a human APC protein which is the product of a mutant allele found in a tumor, wherein the antibodies do not specifically bind to other human proteins, and wherein the human APC protein is a mutant form of the amino acid sequence shown in SEQ ID NOS:2 and 7, and the mutant allele is a mutant form of the nucleotide sequence shown in SEQ ID NO:1.

3. The preparation of claim 2 wherein the mutant allele contains a mutation selected from the group consisting of

mutations at codons 243, 279, 288, 301,331,413,437, 456, 500, 712, and 1338.

4. The preparation of claim 2 wherein the mutant allele

contains a premature stop codon.

5. The preparation of claim 2 wherein the mutant allele contains a missense mutation.

6. The preparation of claim 2 wherein the mutant allele

contains a frameshift mutation.

7. The preparation of claim 2 wherein the mutant allele contains a splice junction mutation.

45

8. The preparation of claim 2 wherein the mutant allele contains an insertion mutation.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: ALBERTSEN, HANS
 ANAND, RAKESH
 CARLSON, MARY
 GRODEN, JOANNA
 HEDGE, PHILIP J.
 JOSLYN, GEOFF
 KINZLER, KENNETH
 MARKHAM, ALEXANDER F.
 NAKAMURA, YUSUKE
 THLIVERIS, ANDREW
 VOGELSTEIN, BERT
 WHITE, RAYMOND L.
- (ii) TITLE OF INVENTION: APC ANTIBODIES
- (iii) NUMBER OF SEQUENCES: 102
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Banner & Witcoff, Ltd.
 - (B) STREET: 1001 G Street, NW
 - (C) CITY: Washington
 - (D) STATE: D.C.
 - (E) COUNTRY: USA
 - (F) ZIP: 20001-4598
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
 - (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/452,654
 - (B) FILING DATE: 25-MAY-1995
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/289,548
 - (B) FILING DATE: 12-AUG-1994
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 07/741,940
 - (B) FILING DATE: 08-AUG-1991
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Kagan, Sarah A.
 - (B) REGISTRATION NUMBER: 32,141

	(ix		LECON A) TE B) TE	ELEPI	ONE	202	2-508	3-910						
	(2) INF	ORMAT	TION	FOR	SEQ	ID 1	NO:1	:						
	(i	(I (C	QUENCA) LE 3) TY C) SY O) TO	ENGTI (PE : [RANI	H: 96 nucl	506 l leic ESS:	oase acio doul	pai:	cs					
	(ii) MOI	LECUI	E T	PE:	cDN	£							
	(vi) OR:	IGINA A) OF				sag	piens	5					
	(vii	IMI (MEDI <i>A</i> 3) CI				APC)							
Had Verd I B Gree S B B B Brit York York	(ix		ATURE A) NA 3) LO	ME/I			. 8562	2						
	(xi) SE(QUENC	CE DI	ESCR.	IPTIC	ON: S	SEQ I	ID NO	0:1:				
fand derd	GGACTCG	GAA A	ATGAC	EGTC	CA AC	GGT <i>I</i>	AGCC?	A AGO	Met			a Ser	T GAT c Asp	54
	CAG TTG Gln Leu													102
	CGA CAA Arg Gln 25	Glu												150
	GAG GCA Glu Ala 40													198
	GAA GAT Glu Asp												 	246
	CTT AAA Leu Lys													294
	CGG TCA Arg Ser													342

(C) REFERENCE/DOCKET NUMBER: 1107.78817

90 95 100

TC <i>P</i> Ser	A AGO Ser 105	Arc	TCT J Ser	GGA Gly	GAG Glu	TGC Cys 110	Ser	CCI Pro	GTT Val	CCI Pro	ATO Met	: Gl	TCA Sei	A TTT	CCA Pro	390
AGA Arg 120	Arg	GGG GGG	TTT Phe	GTA Val	AAT Asn 125	. Gly	AGC Ser	AGA Arg	GAA Glu	AGT Ser 130	Thr	GGZ Gly	TAT	TTA	A GAA 1 Glu 135	438
GAA Glu	CTI Leu	'GAG	AAA Lys	GAG Glu 140	Arg	TCA Ser	TTG Leu	CTI Leu	CTT Leu 145	Ala	'GAT Asp	CTI Leu	GAC Asp	AAA Lys 150	GAA Glu	486
GAA Glu	AAG Lys	GAA Glu	AAA Lys 155	GAC Asp	TGG Trp	TAT Tyr	TAC Tyr	GCT Ala 160	Gln	. CTT . Leu	CAG Gln	AAT Asn	CTC Leu 165	Thr	'AAA 'Lys	534
AGA Arg	ATA Ile	GAT Asp 170	Ser	CTT Leu	CCT Pro	TTA Leu	ACT Thr 175	Glu	AAT Asn	TTT Phe	TCC Ser	TTA Leu 180	Gln	ACA Thr	GAT Asp	582
TTG Leu	ACC Thr 185	AGA Arg	AGG Arg	CAA Gln	TTG Leu	GAA Glu 190	TAT Tyr	GAA Glu	GCA Ala	AGG Arg	CAA Gln 195	ATC Ile	AGA Arg	GTT Val	GCG Ala	630
ATG Met 200	GAA Glu	GAA Glu	CAA Gln	CTA Leu	GGT Gly 205	ACC Thr	TGC Cys	CAG Gln	GAT Asp	ATG Met 210	GAA Glu	AAA Lys	CGA Arg	GCA Ala	CAG Gln 215	678
CGA Arg	AGA Arg	ATA Ile	GCC Ala	AGA Arg 220	ATT Ile	CAG Gln	CAA Gln	ATC Ile	GAA Glu 225	AAG Lys	GAC Asp	ATA Ile	CTT Leu	CGT Arg 230	ATA Ile	726
CGA Arg	CAG Gln	CTT Leu	TTA Leu 235	CAG Gln	TCC Ser	CAA Gln	GCA Ala	ACA Thr 240	GAA Glu	GCA Ala	GAG Glu	AGG Arg	TCA Ser 245	TCT Ser	CAG Gln	774
AAC Asn	AAG Lys	CAT His 250	GAA Glu	ACC Thr	GGC Gly	TCA Ser	CAT His 255	GAT Asp	GCT Ala	GAG Glu	CGG Arg	CAG Gln 260	AAT Asn	GAA Glu	GGT Gly	822
CAA Gln	GGA Gly 265	GTG Val	GGA Gly	GAA Glu	ATC Ile	AAC Asn 270	ATG Met	GCA Ala	ACT Thr	TCT Ser	GGT Gly 275	AAT Asn	GGT Gly	CAG Gln	GGT Gly	870
TCA Ser 280	ACT Thr	ACA Thr	CGA Arg	ATG Met	GAC Asp 285	CAT His	GAA Glu	ACA Thr	GCC Ala	AGT Ser 290	GTT Val	TTG Leu	AGT Ser	TCT Ser	AGT Ser 295	918
AGC Ser	ACA Thr	CAC His	TCT Ser	GCA Ala 300	CCT Pro	CGA Arg	AGG Arg	CTG Leu	ACA Thr 305	AGT Ser	CAT His	CTG Leu	GGA Gly	ACC Thr 310	AAG Lys	966
GTG	GAA	ATG	GTG	TAT	TCA	TTG	TTG	TCA	ATG	CTT	GGT	ACT	CAT	GAT	AAG	1014

Val	Glu	Met	Val 315	Tyr	Ser	Leu	Leu	Ser 320	Met	Leu	Gly	Thr	His 325	Asp	Lys	
					ACT Thr											1062
					CAG Gln											1110
					AAA Lys 365								_			1158
					GCC Ala											1206
					GAC Asp											1254
					ATA Ile											1302
					CCA Pro											1350
					CAG Gln 445											1398
					GAG Glu											1446
					GAA Glu											1494
					TAC Tyr											1542
					ACT Thr											1590
					TGC Cys 525											1638

GAZ Glu	A AGT 1 Ser	GA/	A GAC 1 Asp	TTA Leu 540	Gln	CAG Gln	GTT Val	'ATT	GCA Ala 545	Ser	GTT Val	TTG L Leu	AGG Arg	AAT Asn 550	' TTG . Leu	1686
TCT Ser	TGG Trp	CGA Arg	GCA Ala 555	Asp	GTA Val	AAT Asn	AGT Ser	AAA Lys 560	Lys	ACG Thr	TTG	G CGA 1 Arg	GAA Glu 565	GTT Val	GGA Gly	1734
AGT Ser	GTG Val	Lys 570	Ala	. TTG Leu	ATG Met	GAA Glu	TGT Cys 575	GCT Ala	TTA Leu	GAA Glu	GTT Val	AAA Lys 580	AAG Lys	GAA Glu	TCA Ser	1782
ACC Thr	CTC Leu 585	AAA Lys	AGC Ser	GTA Val	TTG Leu	AGT Ser 590	GCC Ala	TTA Leu	TGG Trp	AAT Asn	TTG Leu 595	TCA Ser	GCA Ala	CAT His	TGC Cys	1830
ACT Thr 600	GIu	AAT Asn	AAA Lys	GCT Ala	GAT Asp 605	ATA Ile	TGT Cys	GCT Ala	GTA Val	GAT Asp 610	GGT Gly	GCA Ala	CTT Leu	GCA Ala	TTT Phe 615	1878
TTG Leu	GTT Val	GGC Gly	ACT Thr	CTT Leu 620	ACT Thr	TAC Tyr	CGG Arg	AGC Ser	CAG Gln 625	ACA Thr	AAC Asn	ACT Thr	TTA Leu	GCC Ala 630	ATT Ile	1926
lle	Glu	Ser	G1y 635	Gly	Gly	Ile	Leu	Arg 640	Asn	Val	Ser	AGC Ser	Leu 645	Ile	Ala	1974
Thr	Asn	G1u 650	Asp	His	Arg	Gln	Ile 655	Leu	Arg	Glu	Asn	AAC Asn 660	Cys	Leu	Gln	2022
Thr	ьеи 665	Leu	GIn	His	Leu	Lys 670	Ser	His	Ser	Leu	Thr 675	ATA Ile	Val	Ser	Asn	2070
680	Cys	GIY	Thr	Leu	Trp 685	Asn	Leu	Ser	Ala	Arg 690	Asn	CCT Pro	Lys	Asp	Gln 695	2118
GIU	Ala	Leu	Trp	700	Met	Gly	Ala	Val	Ser 705	Met	Leu	AAG Lys	Asn	Leu 710	Ile	2166
CAT His	TCA Ser	AAG Lys	CAC His 715	AAA Lys	ATG Met	ATT Ile	GCT Ala	ATG Met 720	GGA Gly	AGT Ser	GCT Ala	GCA Ala	GCT Ala 725	TTA Leu	AGG Arg	2214
Asn	Leu	Met 730	Ala	Asn	Arg	Pro .	Ala 735	Lys	Tyr	Lys	Asp	GCC Ala 740	Asn	Ile	Met	2262
TCT Ser	CCT Pro 745	GGC Gly	TCA Ser	AGC Ser	Leu	CCA Pro 750	TCT Ser	CTT Leu	CAT His	Val	AGG Arg 755	AAA Lys	CAA . Gln	AAA Lys	GCC Ala	2310

CTA Leu 760	ı Glu	GCA Ala	GAA Glu	TTA Leu	GAT Asp 765	Ala	CAG Gln	CAC His	TTA Leu	TCA Ser	Glu	ACT Thr	TTT Phe	GAC Asp	AAT Asn 775	2358
ATA Ile	GAC Asp	: AAT Asn	TTA Leu	AGT Ser 780	Pro	AAG Lys	GCA Ala	TCT Ser	CAT His 785	Arg	AGT Ser	AAG Lys	CAG Gln	AGA Arg 790	CAC His	2406
AAG Lys	CAA Gln	AGT Ser	CTC Leu 795	Tyr	GGT Gly	GAT Asp	TAT Tyr	GTT Val 800	Phe	GAC Asp	ACC Thr	AAT Asn	CGA Arg 805	CAT His	GAT Asp	2454
GAT Asp	AAT Asn	AGG Arg 810	Ser	GAC Asp	AAT Asn	TTT Phe	AAT Asn 815	ACT Thr	GGC Gly	AAC Asn	ATG Met	ACT Thr 820	GTC Val	CTT Leu	TCA Ser	2502
CCA Pro	TAT Tyr 825	Leu	AAT Asn	ACT Thr	ACA Thr	GTG Val 830	TTA Leu	CCC Pro	AGC Ser	TCC Ser	TCT Ser 835	TCA Ser	TCA Ser	AGA Arg	GGA Gly	2550
AGC Ser 840	TTA Leu	GAT Asp	AGT Ser	TCT Ser	CGT Arg 845	TCT Ser	GAA Glu	AAA Lys	GAT Asp	AGA Arg 850	AGT Ser	TTG Leu	GAG Glu	AGA Arg	GAA Glu 855	2598
Arg	Gly	Ile	Gly	Leu 860	Gly	Asn	Tyr	His	Pro 865	Ala	Thr	GAA Glu	Asn	Pro 870	Gly	2646
Thr	Ser	Ser	Lys 875	Arg	Gly	Leu	Gln	Ile 880	Ser	Thr	Thr	GCA Ala	Ala 885	Gln	Ile	2694
Ala	Lys	Val 890	Met	Glu	Glu	Val	Ser 895	Ala	Ile	His	Thr	TCT Ser 900	Gln	Glu	Asp	2742
Arg	Ser 905	Ser	Gly	Ser	Thr	Thr 910	Glu	Leu	His	Cys	Val 915	ACA Thr	Asp	Glu	Arg	2790
Asn 920	Ala	Leu	Arg	Arg	Ser 925	Ser	Ala	Ala	His	Thr 930	His	TCA Ser	Asn	Thr	Tyr 935	2838
AAT Asn	TTC Phe	ACT Thr	AAG Lys	TCG Ser 940	GAA Glu	AAT Asn	TCA Ser	AAT Asn	AGG Arg 945	ACA Thr	TGT Cys	TCT Ser	ATG Met	CCT Pro 950	TAT Tyr	2886
Ala	Lys	Leu	Glu 955	Tyr	Lys	Arg	Ser	Ser 960	Asn	Asp	Ser	TTA Leu	Asn 965	Ser	Val	2934
AGT Ser	AGT Ser	AAT Asn 970	GAT Asp	GGT Gly	TAT Tyr	Gly	AAA Lys 975	AGA Arg	GGT Gly	CAA Gln	ATG Met	AAA Lys 980	CCC Pro	TCG Ser	ATT Ile	2982

GAA Glu	TCC Ser 985	Tyr	TCT Ser	GAA Glu	GAT Asp	GAT Asp 990	Glu	AGT Ser	'AAG	TTT Phe	TGC Cys	Ser	TAT Tyr	GGT Gly	CAA Gln	3030
TAC Tyr 100	Pro	GCC Ala	GAC Asp	CTA Leu	GCC Ala 100	His	AAA Lys	ATA Ile	CAT His	AGT Ser 101	Ala	AAT Asn	CAT His	ATG Met	GAT Asp 1015	3078
GAT Asp	AAT Asn	GAT Asp	GGA Gly	GAA Glu 102	Leu	GAT Asp	ACA Thr	CCA Pro	ATA Ile 102	Asn	TAT Tyr	AGT Ser	CTT Leu	AAA Lys 103	TAT Tyr 0	3126
TCA Ser	GAT Asp	GAG Glu	CAG Gln 103	Leu	AAC Asn	TCT Ser	GGA Gly	AGG Arg 104	Gln	AGT Ser	CCT Pro	TCA Ser	CAG Gln 104	Asn	GAA Glu	3174
AGA Arg	TGG Trp	GCA Ala 105	Arg	CCC Pro	AAA Lys	CAC His	ATA Ile 105	Ile	GAA Glu	GAT Asp	GAA Glu	ATA Ile 106	Lys	CAA Gln	AGT Ser	3222
GAG Glu	CAA Gln 106	Arg	CAA Gln	TCA Ser	AGG Arg	AAT Asn 1070	Gln	AGT Ser	ACA Thr	ACT Thr	TAT Tyr 107	Pro	GTT Val	TAT Tyr	ACT Thr	3270
GAG Glu 108	AGC Ser 0	ACT Thr	GAT Asp	GAT Asp	AAA Lys 1085	His	CTC Leu	AAG Lys	TTC Phe	CAA Gln 1090	Pro	CAT His	TTT Phe	GGA Gly	CAG Gln 1095	3318
CAG Gln	GAA Glu	TGT Cys	GTT Val	TCT Ser 1100	Pro	TAC Tyr	AGG Arg	TCA Ser	CGG Arg 110	Gly	GCC Ala	AAT Asn	GGT Gly	TCA Ser 1110	Glu	3366
ACA Thr	AAT Asn	CGA Arg	GTG Val 1119	Gly	TCT Ser	AAT Asn	CAT His	GGA Gly 1120	Ile	AAT Asn	CAA Gln	AAT Asn	GTA Val 1125	Ser	CAG Gln	3414
TCT Ser	TTG Leu	TGT Cys 1130	Gln	GAA Glu	GAT Asp	GAC Asp	TAT Tyr 1135	Glu	GAT Asp	GAT Asp	AAG Lys	CCT Pro 1140	Thr	AAT Asn	TAT Tyr	3462
AGT Ser	GAA Glu 1145	Arg	TAC Tyr	TCT Ser	GAA Glu	GAA Glu 1150	Glu	CAG Gln	CAT His	GAA Glu	GAA Glu 1155	Glu	GAG Glu	AGA Arg	CCA Pro	3510
ACA Thr 1160	AAT Asn	TAT Tyr	AGC Ser	ATA Ile	AAA Lys 1165	Tyr	AAT Asn	GAA Glu	GAG Glu	AAA Lys 1170	Arg	CAT His	GTG Val	GAT Asp	CAG Gln 1175	3558
CCT Pro	ATT Ile	GAT Asp	TAT Tyr	AGT Ser 1180	Leu	AAA Lys	TAT Tyr	GCC Ala	ACA Thr 1185	Asp	ATT Ile	CCT Pro	TCA Ser	TCA Ser 1190	Gln	3606
AAA Lys	CAG Gln	TCA Ser	TTT Phe 1195	Ser	TTC Phe	TCA :	Lys	AGT Ser 1200	Ser	TCT Ser	GGA Gly	CAA Gln	AGC Ser 1205	Ser	AAA Lys	3654

ACC Thr	GAA Glu	CAT His 121	Met	TCT Ser	TCA Ser	AGC Ser	AGT Ser 121	Glu	AAT Asn	ACG Thr	TCC Ser	ACA Thr	Pro	TCA Ser	TCT Ser	3702
AAT Asn	GCC Ala 122	Lys	AGG Arg	CAG Gln	AAT Asn	CAG Gln 123	Leu	CAT His	CCA Pro	AGT Ser	TCT Ser 123	Ala	CAG Gln	AGT Ser	AGA Arg	3750
AGT Ser 1240	Gly	CAG Gln	CCT Pro	CAA Gln	AAG Lys 124	Ala	GCC Ala	ACT Thr	TGC Cys	AAA Lys 125	Val	TCT Ser	TCT Ser	ATT Ile	AAC Asn 1255	3798
CAA Gln	GAA Glu	ACA Thr	ATA Ile	CAG Gln 126	Thr	TAT Tyr	TGT Cys	GTA Val	GAA Glu 126	Asp	ACT Thr	CCA Pro	ATA Ile	TGT Cys 127	Phe	3846
TCA Ser	AGA Arg	TGT Cys	AGT Ser 127	Ser	TTA Leu	TCA Ser	TCT Ser	TTG Leu 128	Ser	TCA Ser	GCT Ala	GAA Glu	GAT Asp 128	Glu	ATA Ile	3894
GGA Gly	TGT Cys	AAT Asn 129	Gln	ACG Thr	ACA Thr	CAG Gln	GAA Glu 129	Ala	GAT Asp	TCT Ser	GCT Ala	AAT Asn 130	Thr	CTG Leu	CAA Gln	3942
ATA Ile	GCA Ala 1305	Glu	ATA Ile	AAA Lys	GGA Gly	AAG Lys 131	Ile	GGA Gly	ACT Thr	AGG Arg	TCA Ser 131	GCT Ala 5	GAA Glu	GAT Asp	CCT Pro	3990
GTG Val 1320	Ser	GAA Glu	GTT Val	CCA Pro	GCA Ala 1325	Val	TCA Ser	CAG Gln	CAC His	CCT Pro 1330	Arg	ACC Thr	AAA Lys	TCC Ser	AGC Ser 1335	4038
AGA Arg	CTG Leu	CAG Gln	GGT Gly	TCT Ser 1340	Ser	TTA Leu	TCT Ser	TCA Ser	GAA Glu 1345	Ser	GCC Ala	AGG Arg	CAC His	AAA Lys 1350	Ala	4086
GTT Val	GAA Glu	TTT Phe	CCT Pro 1355	Ser	GGA Gly	GCG Ala	AAA Lys	TCT Ser 1360	Pro	TCC Ser	AAA Lys	AGT Ser	GGT Gly 1365	Ala	CAG Gln	4134
ACA Thr	CCC Pro	AAA Lys 1370	Ser	CCA Pro	CCT Pro	GAA Glu	CAC His 1375	Tyr	GTT Val	CAG Gln	GAG Glu	ACC Thr 1380	Pro	CTC Leu	ATG Met	4182
Phe	AGC Ser 1385	Arg	TGT Cys	ACT Thr	TCT Ser	GTC Val 1390	Ser	TCA Ser	CTT Leu	GAT Asp	AGT Ser 1395		GAG Glu	AGT Ser	CGT Arg	4230
TCG Ser 1400	Ile	GCC Ala	AGC Ser	Ser	GTT Val 1405	Gln	AGT Ser	GAA Glu	CCA Pro	TGC Cys 1410	Ser	GGA Gly	ATG Met	GTA Val	AGT Ser 1415	4278
GGC . Gly	ATT Ile	ATA Ile	AGC Ser	CCC Pro 1420	Ser	GAT Asp	CTT Leu	CCA Pro	GAT Asp 1425	Ser	CCT Pro	GGA Gly	CAA Gln	ACC Thr 1430	ATG Met	4326

CCA Pro	CCA Pro	AGC Ser	AGA Arg 143	, Ser	AAA Lys	ACA Thr	CCI Pro	CCA Pro 144	Pro	A CCT	CCT Pro	CAA Glr	A ACA 1 Thi 144	: Ala	r CAA a Gln	4374
ACC Thr	AAG Lys	CGA Arg 145	Glu	GTA Val	CCT Pro	' AAA Lys	AAT Asn 145	Lys	GCA Ala	A CCI Pro	ACT Thr	GCT Ala 146	Glu	A AAG 1 Lys	G AGA G Arg	4422
GAG Glu	AGT Ser 146	GLy	CCT Pro	AAG Lys	Gln	GCT Ala 147	Ala	GTA Val	AAT Asn	GCT Ala	GCA Ala 147	. Val	CAG	AGG Arg	GTC Val	4470
CAG Gln 1480	Val	CTT Leu	CCA Pro	GAT Asp	GCT Ala 148	Asp	ACT Thr	TTA Leu	TTA Leu	CAT His 149	Phe	'GCC Ala	ACA Thr	GAA Glu	AGT Ser 1495	4518
ACT Thr	CCA Pro	GAT Asp	GGA Gly	TTT Phe 150	Ser	TGT Cys	TCA Ser	TCC Ser	AGC Ser 150	Leu	AGT Ser	GCT Ala	CTG Leu	AGC Ser 151	CTC Leu 0	4566
GAT Asp	GAG Glu	CCA Pro	TTT Phe 151	Ile	CAG Gln	AAA Lys	GAT Asp	GTG Val 152	Glu	TTA Leu	AGA Arg	ATA Ile	ATG Met 152	Pro	CCA Pro	4614
GTT Val	CAG Gln	GAA Glu 153	Asn	GAC Asp	AAT Asn	GGG Gly	AAT Asn 153!	Glu	ACA Thr	GAA Glu	TCA Ser	GAG Glu 154	Gln	CCT Pro	AAA Lys	4662
Glu	TCA Ser 1545	Asn	GAA Glu	AAC Asn	CAA Gln	GAG Glu 1550	Lys	GAG Glu	GCA Ala	GAA Glu	AAA Lys 155!	Thr	ATT Ile	GAT Asp	TCT Ser	4710
GAA Glu 1560	Lys	GAC Asp	CTA Leu	TTA Leu	GAT Asp 1565	Asp	TCA Ser	GAT Asp	GAT Asp	GAT Asp 1570	Asp	ATT Ile	GAA Glu	ATA Ile	CTA Leu 1575	4758
GAA Glu	GAA Glu	TGT Cys	ATT Ile	ATT Ile 1580	Ser	GCC Ala	ATG Met	CCA Pro	ACA Thr 1589	Lys	TCA Ser	TCA Ser	CGT Arg	AAA Lys 1590	Gly	4806
AAA .	AAG Lys	CCA Pro	GCC Ala 1595	Gin	ACT Thr	GCT Ala	TCA Ser	AAA Lys 1600	Leu	CCT Pro	CCA Pro	CCT Pro	GTG Val 1609	Ala	AGG Arg	4854
AAA (Pro	AGT Ser 1610	Gln	CTG Leu	CCT Pro	Val	TAC Tyr 1615	Lys	CTT Leu	CTA Leu	CCA Pro	TCA Ser 1620	Gln	AAC Asn	AGG Arg	4902
TTG (CAA Gln 1625	CCC Pro	CAA Gln	AAG Lys	CAT His	GTT Val 1630	Ser	TTT Phe	ACA Thr	CCG Pro	GGG Gly 1635	Asp	GAT Asp	ATG Met	CCA Pro	4950
CGG (Arg V 1640	GTG Val	TAT Tyr	TGT Cys	GTT Val	GAA Glu 1645	Gly	ACA Thr	CCT Pro	ATA Ile	AAC Asn 1650	Phe	TCC Ser	ACA Thr	GCT Ala	ACA Thr 1655	4998

TCI Ser	CTA	AGT Ser	GAT Asp	CTA Leu 166	Thr	ATC Ile	GAA Glu	TCC Ser	CCT Pro 166	Pro	AAT Asn	GAG Glu	TTA Leu	GCT Ala 167	Ala	5046
GGA Gly	GAA Glu	GGA Gly	GTT Val 167	Arg	GGA Gly	GGA Gly	GCA Ala	CAG Gln 168	Ser	GGT Gly	' GAA ' Glu	TTT Phe	GAA Glu 168	Lys	CGA Arg	5094
GAT Asp	ACC Thr	ATT Ile 169	Pro	ACA Thr	GAA Glu	GGC Gly	AGA Arg 169	Ser	ACA Thr	GAT Asp	GAG Glu	GCT Ala 170	Gln	GGA Gly	GGA Gly	5142
AAA Lys	ACC Thr 170	Ser	TCT Ser	GTA Val	ACC Thr	ATA Ile 171	Pro	GAA Glu	TTG Leu	GAT Asp	GAC Asp 171	Asn	AAA Lys	GCA Ala	GAG Glu	5190
GAA Glu 172	Gly	GAT Asp	ATT Ile	CTT Leu	GCA Ala 172	Glu	TGC Cys	ATT Ile	AAT Asn	TCT Ser 173	GCT Ala 0	ATG Met	CCC Pro	AAA Lys	GGG Gly 1735	5238
AAA Lys	AGT Ser	CAC His	AAG Lys	CCT Pro 174	Phe	CGT Arg	GTG Val	AAA Lys	AAG Lys 174!	Ile	ATG Met	GAC Asp	CAG Gln	GTC Val 175	Gln	5286
CAA Gln	GCA Ala	TCT Ser	GCG Ala 175	Ser	TCT Ser	TCT Ser	GCA Ala	CCC Pro 1760	Asn	AAA Lys	AAT Asn	CAG Gln	TTA Leu 176!	Asp	GGT Gly	5334
AAG Lys	AAA Lys	AAG Lys 177	Lys	CCA Pro	ACT Thr	TCA Ser	CCA Pro 1775	Val	AAA Lys	CCT Pro	ATA Ile	CCA Pro 1780	Gln	AAT Asn	ACT Thr	5382
GAA Glu	TAT Tyr 1785	Arg	ACA Thr	CGT Arg	GTA Val	AGA Arg 1790	Lys	AAT Asn	GCA Ala	GAC Asp	TCA Ser 1795	Lys	AAT Asn	AAT Asn	TTA Leu	5430
AAT Asn 1800	Ala	GAG Glu	AGA Arg	GTT Val	TTC Phe 1805	Ser	GAC Asp	AAC Asn	AAA Lys	GAT Asp 1810	TCA Ser	AAG Lys	AAA Lys	CAG Gln	AAT Asn 1815	5478
TTG Leu	AAA Lys	AAT Asn	AAT Asn	TCC Ser 1820	Lys	GAC Asp	TTC Phe	AAT Asn	GAT Asp 1825	Lys	CTC Leu	CCA Pro	AAT Asn	AAT Asn 1830	Glu	5526
GAT Asp	AGA Arg	GTC Val	AGA Arg 1835	Gly	AGT Ser	TTT Phe	GCT Ala	TTT Phe 1840	Asp	TCA Ser	CCT Pro	CAT His	CAT His 1845	Tyr	ACG Thr	5574
CCT Pro	ATT Ile	GAA Glu 1850	Gly	ACT Thr	CCT Pro	Tyr	TGT Cys 1855	Phe	TCA Ser	CGA Arg	AAT Asn	GAT Asp 1860	Ser	TTG Leu	AGT Ser	5622
TCT Ser	CTA Leu 1865	Asp	TTT Phe	GAT Asp	Asp	GAT Asp 1870	GAT Asp	GTT Val	GAC Asp	CTT Leu	TCC Ser 1875	Arg	GAA Glu	AAG Lys	GCT Ala	5670

GAA Glu 188	ı Lev	AGA ı Arg	AAG Lys	GCA Ala	AAA Lys 188	Glu	AAT Asn	' AAG Lys	GAA Glu	TCA Ser 189	Glu	GCT Ala	AAA Lys	GTI Val	ACC Thr 1895	5718
AGC Ser	CAC His	ACA Thr	GAA Glu	CTA Leu 190	Thr	TCC Ser	AAC Asn	CAA Gln	CAA Gln 190	Ser	GCT Ala	' AAT . Asn	AAG Lys	ACA Thr 191	CAA Gln O	5766
GCT Ala	'ATT	'GCA Ala	AAG Lys 191	Gln	CCA Pro	ATA	AAT Asn	CGA Arg 192	Gly	CAG Gln	CCT Pro	AAA Lys	CCC Pro 192	Ile	CTT Leu	5814
CAG Gln	AAA Lys	CAA Gln 193	Ser	ACT Thr	TTT Phe	CCC Pro	CAG Gln 193	Ser	TCC Ser	AAA Lys	GAC Asp	ATA Ile 194	Pro	GAC Asp	AGA Arg	5862
GGG	GCA Ala 194	GCA Ala 5	ACT Thr	GAT Asp	GAA Glu	AAG Lys 1950	Leu	CAG Gln	AAT Asn	TTT Phe	GCT Ala 195	Ile	GAA Glu	AAT Asn	ACT Thr	5910
CCA Pro 196	Val	TGC Cys	TTT Phe	TCT Ser	CAT His 196	Asn	TCC Ser	TCT Ser	CTG Leu	AGT Ser 197	Ser	CTC Leu	AGT Ser	GAC Asp	ATT Ile 1975	5958
GAC Asp	CAA Gln	GAA Glu	AAC Asn	AAC Asn 1980	Asn	AAA Lys	GAA Glu	AAT Asn	GAA Glu 198	Pro	ATC Ile	AAA Lys	GAG Glu	ACT Thr 199	Glu	6006
CCC Pro	CCT Pro	GAC Asp	TCA Ser 1995	Gln	GGA Gly	GAA Glu	CCA Pro	AGT Ser 2000	Lys	CCT Pro	CAA Gln	GCA Ala	TCA Ser 2009	Gly	TAT Tyr	6054
GCT Ala	CCT Pro	AAA Lys 2010	Ser	TTT Phe	CAT His	GTT Val	GAA Glu 2015	Asp	ACC Thr	CCA Pro	GTT Val	TGT Cys 2020	Phe	TCA Ser	AGA Arg	6102
AAC Asn	AGT Ser 2025	TCT Ser	CTC Leu	AGT Ser	TCT Ser	CTT Leu 2030	Ser	ATT Ile	GAC Asp	TCT Ser	GAA Glu 2035	Asp	GAC Asp	CTG Leu	TTG Leu	6150
CAG Gln 2040	Glu	TGT Cys	ATA Ile	AGC Ser	TCC Ser 2045	Ala	ATG Met	CCA Pro	AAA Lys	AAG Lys 2050	Lys	AAG Lys	CCT Pro	TCA Ser	AGA Arg 2055	6198
CTC Leu	AAG Lys	GGT Gly	Asp	AAT Asn 2060	Glu	AAA Lys	CAT His	AGT Ser	CCC Pro 2065	Arg	AAT Asn	ATG Met	GGT Gly	GGC Gly 2070	Ile	6246
TTA Leu	GGT Gly	GAA Glu	GAT Asp 2075	Leu	ACA Thr	CTT Leu	Asp	TTG Leu 2080	Lys	GAT Asp	ATA Ile	CAG Gln	AGA Arg 2085	Pro	GAT Asp	6294
TCA Ser	GAA Glu	CAT His 2090	Gly	CTA Leu	TCC Ser	Pro	GAT Asp 2095	Ser	GAA Glu	AAT Asn	TTT Phe	GAT Asp 2100	Trp	AAA Lys	GCT Ala	6342

ATT Ile	CAG Gln 210	Glu	GGT Gly	GCA Ala	AAT Asn	TCC Ser 211	Ile	GTA Val	AGT Ser	AGT Ser	TTA Leu 211	His	CAA Gln	GCT Ala	GCT Ala	6390
GCT Ala 212	Ala	GCA Ala	TGT Cys	TTA Leu	TCT Ser 212	Arg	CAA Gln	GCT Ala	TCG Ser	TCT Ser 213	Asp	TCA Ser	GAT Asp	TCC Ser	ATC Ile 2135	6438
			AAA Lys		Gly					Ser						6486
CCT Pro	GAT Asp	CAA Gln	GAA Glu 215	Glu	AAA Lys	CCC Pro	TTT Phe	ACA Thr 216	Ser	AAT Asn	AAA Lys	GGC Gly	CCA Pro 216	Arg	ATT Ile	6534
CTA Leu	AAA Lys	CCA Pro 217	GGG Gly O	GAG Glu	AAA Lys	AGT Ser	ACA Thr 217	Leu	GAA Glu	ACT Thr	AAA Lys	AAG Lys 218	Ile	GAA Glu	TCT Ser	6582
		Lys	GGA Gly				Gly					Lys				6630
ACT Thr 220	Gly	AAA Lys	GTT Val	CGA Arg	TCT Ser 2205	Asn	TCA Ser	GAA Glu	ATT Ile	TCA Ser 221	Gly	CAA Gln	ATG Met	AAA Lys	CAG Gln 2215	6678
			GCA Ala		Met					Arg					Ile	6726
			GGA Gly 2235	Val					Ser					Val		6774
AAA Lys	AAA Lys	GGC Gly 225	CCA Pro	CCC Pro	CTT Leu	AAG Lys	ACT Thr 2255	Pro	GCC Ala	TCC Ser	AAA Lys	AGC Ser 2260	Pro	AGT Ser	GAA Glu	6822
GGT Gly	CAA Gln 2265	Thr	GCC Ala	ACC Thr	ACT Thr	TCT Ser 2270	Pro	AGA Arg	GGA Gly	GCC Ala	AAG Lys 2275	Pro	TCT Ser	GTG Val	AAA Lys	6870
	Glu		AGC Ser			Ala					Gln					6918
AGT Ser	AAA Lys	GCA Ala	CCT Pro	TCT Ser 2300	Arg	TCA Ser	GGA Gly	TCT Ser	AGA Arg 2305	Asp	TCG Ser	ACC Thr	CCT Pro	TCA Ser 2310	Arg	6966
			CAA Gln 2315	Pro					Ile					Arg		7014

TCA Ser	ATT Ile	TCC Ser 233	Pro	GGT Gly	AGA Arg	AAT Asn	GGA Gly 233	Ile	. AGT Ser	CCT Pro	CCT Pro	AAC Asn 234	Lys	TTA Leu	TCT Ser	7062
CAA Gln	CTT Leu 2345	Pro	AGG Arg	ACA Thr	TCA Ser	TCC Ser 235	Pro	AGT Ser	ACT Thr	GCT Ala	TCA Ser 235	Thr	AAG Lys	TCC Ser	TCA Ser	7110
GGT Gly 236	Ser	GGA Gly	AAA Lys	ATG Met	TCA Ser 236	Tyr	ACA Thr	TCT Ser	CCA Pro	GGT Gly 237	Arg	CAG Gln	ATG Met	AGC Ser	CAA Gln 2375	7158
CAG Gln	AAC Asn	CTT Leu	ACC Thr	AAA Lys 238	Gln	ACA Thr	GGT Gly	TTA Leu	TCC Ser 238	Lys	AAT Asn	GCC Ala	AGT Ser	AGT Ser 239	Ile	7206
CCA Pro	AGA Arg	AGT Ser	GAG Glu 239	Ser	GCC Ala	TCC Ser	AAA Lys	GGA Gly 240	Leu	AAT Asn	CAG Gln	ATG Met	AAT Asn 240	Asn	GGT Gly	7254
AAT Asn	GGA Gly	GCC Ala 241	Asn	AAA Lys	AAG Lys	GTA Val	GAA Glu 241	Leu	TCT Ser	AGA Arg	ATG Met	TCT Ser 242	Ser	ACT Thr	AAA Lys	7302
TCA Ser	AGT Ser 2425	Gly	AGT Ser	GAA Glu	TCT Ser	GAT Asp 2430	Arg	TCA Ser	GAA Glu	AGA Arg	CCT Pro 243	Val	TTA Leu	GTA Val	CGC Arg	7350
CAG Gln 244(TCA Ser	ACT Thr	TTC Phe	ATC Ile	AAA Lys 244!	Glu	GCT Ala	CCA Pro	AGC Ser	CCA Pro 2450	Thr	TTA Leu	AGA Arg	AGA Arg	AAA Lys 2455	7398
TTG Leu	GAG Glu	GAA Glu	TCT Ser	GCT Ala 2460	Ser	TTT Phe	GAA Glu	TCT Ser	CTT Leu 2465	Ser	CCA Pro	TCA Ser	TCT Ser	AGA Arg 2470	Pro	7446
GCT Ala	TCT Ser	CCC Pro	ACT Thr 2475	Arg	TCC Ser	CAG Gln	GCA Ala	CAA Gln 2480	Thr	CCA Pro	GTT Val	TTA Leu	AGT Ser 248	Pro	TCC Ser	7494
	CCT Pro		Met					His					Ala			7542
	CGA Arg 2505	Lys					Leu					Glu				7590
GGA Gly 2520	AGA Arg	CCA Pro	GCA Ala	AAG Lys	CGC Arg 2525	His	GAT Asp	ATT Ile	GCA Ala	CGG Arg 2530	Ser	CAT His	TCT Ser	GAA Glu	AGT Ser 2535	7638
CCT Pro	TCT . Ser .	AGA Arg	Leu	CCA Pro 2540	Ile	AAT Asn	AGG Arg	TCA Ser	GGA Gly 2545	Thr	TGG Trp	AAA Lys	CGT Arg	GAG Glu 2550	His	7686

AGC Ser	AAA Lys	CAT His	TCA Ser 255	Ser	TCC Ser	CTT Leu	CCT Pro	CGA Arg 256	Val	AGC Ser	ACT Thr	TGG Trp	AGA Arg 256	Arg	ACT	7734
GGA Gly	AGT Ser	TCA Ser 257	Ser	TCA Ser	ATT Ile	CTT Leu	TCT Ser 257	Ala	TCA Ser	TCA Ser	GAA Glu	TCC Ser 258	Ser	GAA Glu	AAA Lys	7782
GCA Ala	AAA Lys 258	Ser	GAG Glu	GAT Asp	GAA Glu	AAA Lys 259	His	GTG Val	AAC Asn	TCT Ser	ATT Ile 259	Ser	GGA Gly	ACC Thr	AAA Lys	7830
CAA Gln 260	Ser	AAA Lys	GAA Glu	AAC Asn	CAA Gln 260	Val	TCC Ser	GCA Ala	AAA Lys	GGA Gly 261	Thr	TGG Trp	AGA Arg	AAA Lys	ATA Ile 2615	7878
AAA Lys	GAA Glu	AAT Asn	GAA Glu	TTT Phe 262	Ser	CCC Pro	ACA Thr	AAT Asn	AGT Ser 262	Thr	TCT Ser	CAG Gln	ACC Thr	GTT Val 263	Ser	7926
TCA Ser	GGT Gly	GCT Ala	ACA Thr 263	Asn	GGT Gly	GCT Ala	GAA Glu	TCA Ser 264	Lys	ACT Thr	CTA Leu	ATT Ile	TAT Tyr 264	Gln	ATG Met	7974
GCA Ala	CCT Pro	GCT Ala 265	Val	TCT Ser	AAA Lys	ACA Thr	GAG Glu 265	Asp	GTT Val	TGG Trp	GTG Val	AGA Arg 2660	Ile	GAG Glu	GAC Asp	8022
TGT Cys	CCC Pro 2665	Ile	AAC Asn	AAT Asn	CCT Pro	AGA Arg 2670	Ser	GGA Gly	AGA Arg	TCT Ser	CCC Pro 267	Thr	GGT Gly	AAT Asn	ACT Thr	8070
CCC Pro 2680	Pro	GTG Val	ATT Ile	GAC Asp	AGT Ser 2685	GTT Val	TCA Ser	GAA Glu	AAG Lys	GCA Ala 2690	Asn	CCA Pro	AAC Asn	ATT Ile	AAA Lys 2695	8118
GAT Asp	TCA Ser	AAA Lys	GAT Asp	AAT Asn 2700	Gln	GCA Ala	AAA Lys	CAA Gln	AAT Asn 2705	Val	GGT Gly	AAT Asn	GGC Gly	AGT Ser 2710	Val	8166
CCC Pro	ATG Met	CGT Arg	ACC Thr 2715	Val	GGT Gly	TTG Leu	GAA Glu	AAT Asn 2720	Arg	CTG Leu	ACC Thr	TCC Ser	TTT Phe 2725	Ile	CAG Gln	8214
GTG Val	GAT Asp	GCC Ala 2730	Pro	GAC Asp	CAA Gln	AAA Lys	GGA Gly 2735	Thr	GAG Glu	ATA Ile	AAA Lys	CCA Pro 2740	Gly	CAA Gln	AAT Asn	8262
AAT Asn	CCT Pro 2745	Val	CCT Pro	GTA Val	TCA Ser	GAG Glu 2750	Thr	AAT Asn	GAA Glu	AGT Ser	CCT Pro 2755	Ile	GTG Val	GAA Glu	CGT Arg	8310
ACC Thr 2760	Pro	TTC Phe	AGT Ser	TCT Ser	AGC Ser 2765	AGC Ser	TCA Ser	AGC Ser	AAA Lys	CAC His 2770	Ser	TCA Ser	CCT Pro	AGT Ser	GGG Gly 2775	8358

ACT GTT GCT GCC AGA GTG ACT CCT TTT AAT TAC AAC CCA AGC CCT AGG Thr Val Ala Ala Arg Val Thr Pro Phe Asn Tyr Asn Pro Ser Pro Arg 2780 2785 2790	8406
AAA AGC AGC GCA GAT AGC ACT TCA GCT CGG CCA TCT CAG ATC CCA ACT Lys Ser Ser Ala Asp Ser Thr Ser Ala Arg Pro Ser Gln Ile Pro Thr 2795 2800 2805	8454
CCA GTG AAT AAC AAC ACA AAG AAG CGA GAT TCC AAA ACT GAC AGC ACA Pro Val Asn Asn Asn Thr Lys Lys Arg Asp Ser Lys Thr Asp Ser Thr 2810 2815 2820	8502
GAA TCC AGT GGA ACC CAA AGT CCT AAG CGC CAT TCT GGG TCT TAC CTT Glu Ser Ser Gly Thr Gln Ser Pro Lys Arg His Ser Gly Ser Tyr Leu 2825 2830 2835	8550
GTG ACA TCT GTT TAAAAGAGAG GAAGAATGAA ACTAAGAAAA TTCTATGTTA Val Thr Ser Val 2840	8602
ATTACAACTG CTATATAGAC ATTTTGTTTC AAATGAAACT TTAAAAGACT GAAAAATTTT	8662
GTAAATAGGT TTGATTCTTG TTAGAGGGTT TTTGTTCTGG AAGCCATATT TGATAGTATA	8722
CTTTGTCTTC ACTGGTCTTA TTTTGGGAGG CACTCTTGAT GGTTAGGAAA AAATAGAAAG	8782
CCAAGTATGT TTGTACAGTA TGTTTTACAT GTATTTAAAG TAGCATCCCA TCCCAACTTC	8842
CTTAATTATT GCTTGTCTAA AATAATGAAC ACTACAGATA GGAAATATGA TATATTGCTG	8902
TTATCAATCA TTTCTAGATT ATAAACTGAC TAAACTTACA TCAGGGGAAA ATTGGTATTT	8962
ATGCAAAAA AAAATGTTTT TGTCCTTGTG AGTCCATCTA ACATCATAAT TAATCATGTG	9022
GCTGTGAAAT TCACAGTAAT ATGGTTCCCG ATGAACAAGT TTACCCAGCC TGCTTTGCTT	9082
ACTGCATGAA TGAAACTGAT GGTTCAATTT CAGAAGTAAT GATTAACAGT TATGTGGTCA	9142
CATGATGTGC ATAGAGATAG CTACAGTGTA ATAATTTACA CTATTTTGTG CTCCAAACAA	9202
AACAAAAATC TGTGTAACTG TAAAACATTG AATGAAACTA TTTTACCTGA ACTAGATTTT	9262
ATCTGAAAGT AGGTAGAATT TTTGCTATGC TGTAATTTGT TGTATATTCT GGTATTTGAG	9322
GTGAGATGGC TGCTCTTTAT TAATGAGACA TGAATTGTGT CTCAACAGAA ACTAAATGAA	9382
CATTTCAGAA TAAATTATTG CTGTATGTAA ACTGTTACTG AAATTGGTAT TTGTTTGAAG	9442
GGTTTGTTTC ACATTTGTAT TAATTAATTG TTTAAAATGC CTCTTTTAAA AGCTTATATA	9502
AATTTTTCT TCAGCTTCTA TGCATTAAGA GTAAAATTCC TCTTACTGTA ATAAAAACAT	9562
TGAAGAAGAC TGTTGCCACT TAACCATTCC ATGCGTTGGC ACTT	9606

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2843 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:
- Met Ala Ala Ser Tyr Asp Gln Leu Leu Lys Gln Val Glu Ala Leu

 1 10 15
- Lys Met Glu Asn Ser Asn Leu Arg Gln Glu Leu Glu Asp Asn Ser Asn 20 25 30
- His Leu Thr Lys Leu Glu Thr Glu Ala Ser Asn Met Lys Glu Val Leu 35 40 45
- Lys Gln Leu Gln Gly Ser Ile Glu Asp Glu Ala Met Ala Ser Ser Gly 50 55 60
- Gln Ile Asp Leu Leu Glu Arg Leu Lys Glu Leu Asn Leu Asp Ser Ser 65 70 75 80
- Asn Phe Pro Gly Val Lys Leu Arg Ser Lys Met Ser Leu Arg Ser Tyr 85 90 95
- Gly Ser Arg Glu Gly Ser Val Ser Ser Arg Ser Gly Glu Cys Ser Pro 100 105 110
- Val Pro Met Gly Ser Phe Pro Arg Arg Gly Phe Val Asn Gly Ser Arg 115 120 125
- Glu Ser Thr Gly Tyr Leu Glu Glu Leu Glu Lys Glu Arg Ser Leu Leu 130 135 140
- Leu Ala Asp Leu Asp Lys Glu Glu Lys Glu Lys Asp Trp Tyr Tyr Ala 145 150 155 160
- Gln Leu Gln Asn Leu Thr Lys Arg Ile Asp Ser Leu Pro Leu Thr Glu 165 170 175
- Asn Phe Ser Leu Gln Thr Asp Leu Thr Arg Arg Gln Leu Glu Tyr Glu 180 185 190
- Ala Arg Gln Ile Arg Val Ala Met Glu Glu Gln Leu Gly Thr Cys Gln
 195 200 205
- Asp Met Glu Lys Arg Ala Gln Arg Arg Ile Ala Arg Ile Gln Gln Ile 210 215 220
- Glu Lys Asp Ile Leu Arg Ile Arg Gln Leu Leu Gln Ser Gln Ala Thr 225 230 235 240

- Glu Ala Glu Arg Ser Ser Gln Asn Lys His Glu Thr Gly Ser His Asp 245 250 255
- Ala Glu Arg Gln Asn Glu Gly Gln Gly Val Gly Glu Ile Asn Met Ala 260 265 270
- Thr Ser Gly Asn Gly Gln Gly Ser Thr Thr Arg Met Asp His Glu Thr 275 280 285
- Ala Ser Val Leu Ser Ser Ser Ser Thr His Ser Ala Pro Arg Arg Leu 290 295 300
- Thr Ser His Leu Gly Thr Lys Val Glu Met Val Tyr Ser Leu Leu Ser 305 310 315 320
- Met Leu Gly Thr His Asp Lys Asp Asp Met Ser Arg Thr Leu Leu Ala 325 330 335
- Met Ser Ser Gln Asp Ser Cys Ile Ser Met Arg Gln Ser Gly Cys 340 345 350
- Leu Pro Leu Leu Ile Gln Leu Leu His Gly Asn Asp Lys Asp Ser Val 355 360 365
- Leu Leu Gly Asn Ser Arg Gly Ser Lys Glu Ala Arg Ala Arg Ala Ser 370 375 380
- Ala Ala Leu His Asn Ile Ile His Ser Gln Pro Asp Asp Lys Arg Gly 385 390 395 400
- Arg Arg Glu Ile Arg Val Leu His Leu Leu Glu Gln Ile Arg Ala Tyr 405 410 415
- Cys Glu Thr Cys Trp Glu Trp Gln Glu Ala His Glu Pro Gly Met Asp 420 425 430
- Gln Asp Lys Asn Pro Met Pro Ala Pro Val Glu His Gln Ile Cys Pro 435 440 445
- Ala Val Cys Val Leu Met Lys Leu Ser Phe Asp Glu Glu His Arg His 450 455 460
- Ala Met Asn Glu Leu Gly Gly Leu Gln Ala Ile Ala Glu Leu Leu Gln 465 470 475 480
- Val Asp Cys Glu Met Tyr Gly Leu Thr Asn Asp His Tyr Ser Ile Thr 485 490 495
- Leu Arg Arg Tyr Ala Gly Met Ala Leu Thr Asn Leu Thr Phe Gly Asp 500 505 510
- Val Ala Asn Lys Ala Thr Leu Cys Ser Met Lys Gly Cys Met Arg Ala
 515 520 525
- Leu Val Ala Gln Leu Lys Ser Glu Ser Glu Asp Leu Gln Gln Val Ile

530 535 540

Ala Ser Val Leu Arg Asn Leu Ser Trp Arg Ala Asp Val Asn Ser Lys 545 550 555 560

Lys Thr Leu Arg Glu Val Gly Ser Val Lys Ala Leu Met Glu Cys Ala 565 570 575

Leu Glu Val Lys Lys Glu Ser Thr Leu Lys Ser Val Leu Ser Ala Leu 580 585

Trp Asn Leu Ser Ala His Cys Thr Glu Asn Lys Ala Asp Ile Cys Ala 595 600 605

Val Asp Gly Ala Leu Ala Phe Leu Val Gly Thr Leu Thr Tyr Arg Ser 610 620

Gln Thr Asn Thr Leu Ala Ile Ile Glu Ser Gly Gly Gly Ile Leu Arg 625 630 635 640

Asn Val Ser Ser Leu Ile Ala Thr Asn Glu Asp His Arg Gln Ile Leu 645 650 655

Arg Glu Asn Asn Cys Leu Gln Thr Leu Leu Gln His Leu Lys Ser His
660 665 670

Ser Leu Thr Ile Val Ser Asn Ala Cys Gly Thr Leu Trp Asn Leu Ser 675 680 685

Ala Arg Asn Pro Lys Asp Gln Glu Ala Leu Trp Asp Met Gly Ala Val 690 695 700

Ser Met Leu Lys Asn Leu Ile His Ser Lys His Lys Met Ile Ala Met 705 710 715 720

Gly Ser Ala Ala Ala Leu Arg Asn Leu Met Ala Asn Arg Pro Ala Lys
725 730 735

Tyr Lys Asp Ala Asn Ile Met Ser Pro Gly Ser Ser Leu Pro Ser Leu 740 745 750

His Val Arg Lys Gln Lys Ala Leu Glu Ala Glu Leu Asp Ala Gln His
755 760 765

Leu Ser Glu Thr Phe Asp Asn Ile Asp Asn Leu Ser Pro Lys Ala Ser 770 780

His Arg Ser Lys Gln Arg His Lys Gln Ser Leu Tyr Gly Asp Tyr Val 785 790 795 800

Phe Asp Thr Asn Arg His Asp Asp Asn Arg Ser Asp Asn Phe Asn Thr 805 810 815

Gly Asn Met Thr Val Leu Ser Pro Tyr Leu Asn Thr Thr Val Leu Pro

820 825 830

Ser Ser Ser Ser Ser Arg Gly Ser Leu Asp Ser Ser Arg Ser Glu Lys 835 840 845

Asp Arg Ser Leu Glu Arg Glu Arg Gly Ile Gly Leu Gly Asn Tyr His 850 855 860

Pro Ala Thr Glu Asn Pro Gly Thr Ser Ser Lys Arg Gly Leu Gln Ile 865 870 875 880

Ser Thr Thr Ala Ala Gln Ile Ala Lys Val Met Glu Glu Val Ser Ala 885 890 895

Ile His Thr Ser Gln Glu Asp Arg Ser Ser Gly Ser Thr Thr Glu Leu
900 905 910

His Cys Val Thr Asp Glu Arg Asn Ala Leu Arg Arg Ser Ser Ala Ala 915 920 925

His Thr His Ser Asn Thr Tyr Asn Phe Thr Lys Ser Glu Asn Ser Asn 930 935 940

Arg Thr Cys Ser Met Pro Tyr Ala Lys Leu Glu Tyr Lys Arg Ser Ser 945 950 955 960

Asn Asp Ser Leu Asn Ser Val Ser Ser Asn Asp Gly Tyr Gly Lys Arg 965 970 975

Gly Gln Met Lys Pro Ser Ile Glu Ser Tyr Ser Glu Asp Asp Glu Ser 980 985 990

Lys Phe Cys Ser Tyr Gly Gln Tyr Pro Ala Asp Leu Ala His Lys Ile 995 1000 1005

His Ser Ala Asn His Met Asp Asp Asn Asp Gly Glu Leu Asp Thr Pro 1010 1015 1020

Ile Asn Tyr Ser Leu Lys Tyr Ser Asp Glu Gln Leu Asn Ser Gly Arg 1025 1030 1035 1040

Gln Ser Pro Ser Gln Asn Glu Arg Trp Ala Arg Pro Lys His Ile Ile 1045 1050 1055

Glu Asp Glu Ile Lys Gln Ser Glu Gln Arg Gln Ser Arg Asn Gln Ser 1060 1065 1070

Thr Thr Tyr Pro Val Tyr Thr Glu Ser Thr Asp Asp Lys His Leu Lys 1075 1080 1085

Phe Gln Pro His Phe Gly Gln Gln Glu Cys Val Ser Pro Tyr Arg Ser 1090 1095 1100

Arg Gly Ala Asn Gly Ser Glu Thr Asn Arg Val Gly Ser Asn His Gly

Ile Asn Gln Asn Val Ser Gln Ser Leu Cys Gln Glu Asp Asp Tyr Glu 1125 1130 1135

1115

Asp Asp Lys Pro Thr Asn Tyr Ser Glu Arg Tyr Ser Glu Glu Glu Gln
1140 1145 1150

His Glu Glu Glu Arg Pro Thr Asn Tyr Ser Ile Lys Tyr Asn Glu 1155 1160 1165

Glu Lys Arg His Val Asp Gln Pro Ile Asp Tyr Ser Leu Lys Tyr Ala 1170 1175 1180

Thr Asp Ile Pro Ser Ser Gln Lys Gln Ser Phe Ser Phe Ser Lys Ser 1185 1190 1195 1200

Ser Ser Gly Gln Ser Ser Lys Thr Glu His Met Ser Ser Ser Glu 1205 1210 1215

Asn Thr Ser Thr Pro Ser Ser Asn Ala Lys Arg Gln Asn Gln Leu His 1220 1225 1230

Pro Ser Ser Ala Gln Ser Arg Ser Gly Gln Pro Gln Lys Ala Ala Thr 1235 1240 1245

Cys Lys Val Ser Ser Ile Asn Gln Glu Thr Ile Gln Thr Tyr Cys Val 1250 1255 1260

Glu Asp Thr Pro Ile Cys Phe Ser Arg Cys Ser Ser Leu Ser Ser Leu 1265 1270 1275 1280

Ser Ser Ala Glu Asp Glu Ile Gly Cys Asn Gln Thr Thr Gln Glu Ala 1285 1290 1295

Asp Ser Ala Asn Thr Leu Gln Ile Ala Glu Ile Lys Gly Lys Ile Gly
1300 1310

Thr Arg Ser Ala Glu Asp Pro Val Ser Glu Val Pro Ala Val Ser Gln 1315 1320 1325

His Pro Arg Thr Lys Ser Ser Arg Leu Gln Gly Ser Ser Leu Ser Ser 1330 1335 1340

Glu Ser Ala Arg His Lys Ala Val Glu Phe Pro Ser Gly Ala Lys Ser 1345 1350 1355 1360

Pro Ser Lys Ser Gly Ala Gln Thr Pro Lys Ser Pro Pro Glu His Tyr 1365 1370 1375

Val Gln Glu Thr Pro Leu Met Phe Ser Arg Cys Thr Ser Val Ser Ser 1380 1385 1390

Leu Asp Ser Phe Glu Ser Arg Ser Ile Ala Ser Ser Val Gln Ser Glu

1395 1400 1405

Pro Cys Ser Gly Met Val Ser Gly Ile Ile Ser Pro Ser Asp Leu Pro 1410 1415 1420

Asp Ser Pro Gly Gln Thr Met Pro Pro Ser Arg Ser Lys Thr Pro Pro 1425 1430 1435 1440

Pro Pro Pro Gln Thr Ala Gln Thr Lys Arg Glu Val Pro Lys Asn Lys 1445 1450 1455

Ala Pro Thr Ala Glu Lys Arg Glu Ser Gly Pro Lys Gln Ala Ala Val 1460 1465 1470

Asn Ala Ala Val Gln Arg Val Gln Val Leu Pro Asp Ala Asp Thr Leu 1475 1480 1485

Leu His Phe Ala Thr Glu Ser Thr Pro Asp Gly Phe Ser Cys Ser Ser 1490 1495 1500

Ser Leu Ser Ala Leu Ser Leu Asp Glu Pro Phe Ile Gln Lys Asp Val 1505 1510 1515 1520

Glu Leu Arg Ile Met Pro Pro Val Gln Glu Asn Asp Asn Gly Asn Glu 1525 1530 1535

Thr Glu Ser Glu Gln Pro Lys Glu Ser Asn Glu Asn Gln Glu Lys Glu 1540 1545 1550

Ala Glu Lys Thr Ile Asp Ser Glu Lys Asp Leu Leu Asp Asp Ser Asp 1555 1560 1565

Asp Asp Asp Ile Glu Ile Leu Glu Glu Cys Ile Ile Ser Ala Met Pro 1570 1575 1580.

Thr Lys Ser Ser Arg Lys Gly Lys Lys Pro Ala Gln Thr Ala Ser Lys 1585 1590 1595 1600

Leu Pro Pro Pro Val Ala Arg Lys Pro Ser Gln Leu Pro Val Tyr Lys 1605 1610 1615

Leu Leu Pro Ser Gln Asn Arg Leu Gln Pro Gln Lys His Val Ser Phe 1620 1630

Thr Pro Gly Asp Asp Met Pro Arg Val Tyr Cys Val Glu Gly Thr Pro 1635 1640 1645

Ile Asn Phe Ser Thr Ala Thr Ser Leu Ser Asp Leu Thr Ile Glu Ser 1650 1655 1660

Pro Pro Asn Glu Leu Ala Ala Gly Glu Gly Val Arg Gly Gly Ala Gln 1665 1670 1675 1680

Ser Gly Glu Phe Glu Lys Arg Asp Thr Ile Pro Thr Glu Gly Arg Ser

1685 1690 1695

Thr Asp Glu Ala Gln Gly Gly Lys Thr Ser Ser Val Thr Ile Pro Glu 1700 1710

- Leu Asp Asp Asn Lys Ala Glu Glu Gly Asp Ile Leu Ala Glu Cys Ile 1715 1720 1725
- Asn Ser Ala Met Pro Lys Gly Lys Ser His Lys Pro Phe Arg Val Lys 1730 1735 1740
- Lys Ile Met Asp Gln Val Gln Gln Ala Ser Ala Ser Ser Ser Ala Pro 1745 1750 1755 1760
- Asn Lys Asn Gln Leu Asp Gly Lys Lys Lys Pro Thr Ser Pro Val 1765 1770 1775
- Lys Pro Ile Pro Gln Asn Thr Glu Tyr Arg Thr Arg Val Arg Lys Asn 1780 1785 1790
- Ala Asp Ser Lys Asn Asn Leu Asn Ala Glu Arg Val Phe Ser Asp Asn 1795 1800 1805
- Lys Asp Ser Lys Lys Gln Asn Leu Lys Asn Asn Ser Lys Asp Phe Asn 1810 1815 1820
- Asp Lys Leu Pro Asn Asn Glu Asp Arg Val Arg Gly Ser Phe Ala Phe 1825 1830 1835 1840
- Asp Ser Pro His His Tyr Thr Pro Ile Glu Gly Thr Pro Tyr Cys Phe 1845 1850 1855
- Ser Arg Asn Asp Ser Leu Ser Ser Leu Asp Phe Asp Asp Asp Val 1860 1865 1870
- Asp Leu Ser Arg Glu Lys Ala Glu Leu Arg Lys Ala Lys Glu Asn Lys 1875 1880 1885
- Glu Ser Glu Ala Lys Val Thr Ser His Thr Glu Leu Thr Ser Asn Gln 1890 1895 1900
- Gln Ser Ala Asn Lys Thr Gln Ala Ile Ala Lys Gln Pro Ile Asn Arg 1905 1910 1915 1920
- Gly Gln Pro Lys Pro Ile Leu Gln Lys Gln Ser Thr Phe Pro Gln Ser 1925 1930 1935
- Ser Lys Asp Ile Pro Asp Arg Gly Ala Ala Thr Asp Glu Lys Leu Gln 1940 1945 1950
- Asn Phe Ala Ile Glu Asn Thr Pro Val Cys Phe Ser His Asn Ser Ser 1955 1960 1965
- Leu Ser Ser Leu Ser Asp Ile Asp Gln Glu Asn Asn Asn Lys Glu Asn 1970 1975 1980

- Glu Pro Ile Lys Glu Thr Glu Pro Pro Asp Ser Gln Gly Glu Pro Ser 1985 1990 1995 2000
- Lys Pro Gln Ala Ser Gly Tyr Ala Pro Lys Ser Phe His Val Glu Asp 2005 2010 2015
- Thr Pro Val Cys Phe Ser Arg Asn Ser Ser Leu Ser Ser Leu Ser Ile 2020 2025 2030
- Asp Ser Glu Asp Asp Leu Leu Gln Glu Cys Ile Ser Ser Ala Met Pro 2035 2040 2045
- Lys Lys Lys Pro Ser Arg Leu Lys Gly Asp Asn Glu Lys His Ser 2050 2055 2060
- Pro Arg Asn Met Gly Gly Ile Leu Gly Glu Asp Leu Thr Leu Asp Leu 2065 2070 2075 2080
- Lys Asp Ile Gln Arg Pro Asp Ser Glu His Gly Leu Ser Pro Asp Ser 2085 2090 2095
- Glu Asn Phe Asp Trp Lys Ala Ile Gln Glu Gly Ala Asn Ser Ile Val 2100 2105 2110
- Ser Ser Leu His Gln Ala Ala Ala Ala Cys Leu Ser Arg Gln Ala 2115 2120 2125
- Ser Ser Asp Ser Asp Ser Ile Leu Ser Leu Lys Ser Gly Ile Ser Leu 2130 2135 2140
- Gly Ser Pro Phe His Leu Thr Pro Asp Gln Glu Glu Lys Pro Phe Thr 2145 2150 2155 2160
- Ser Asn Lys Gly Pro Arg Ile Leu Lys Pro Gly Glu Lys Ser Thr Leu 2165 2170 2175
- Glu Thr Lys Lys Ile Glu Ser Glu Ser Lys Gly Ile Lys Gly Gly Lys 2180 2185 2190
- Lys Val Tyr Lys Ser Leu Ile Thr Gly Lys Val Arg Ser Asn Ser Glu 2195 2200 2205
- Ile Ser Gly Gln Met Lys Gln Pro Leu Gln Ala Asn Met Pro Ser Ile 2210 2215 2220
- Ser Arg Gly Arg Thr Met Ile His Ile Pro Gly Val Arg Asn Ser Ser 2225 2230 2235 2240
- Ser Ser Thr Ser Pro Val Ser Lys Lys Gly Pro Pro Leu Lys Thr Pro 2245 2250 2255
- Ala Ser Lys Ser Pro Ser Glu Gly Gln Thr Ala Thr Thr Ser Pro Arg 2260 2265 2270

- Gly Ala Lys Pro Ser Val Lys Ser Glu Leu Ser Pro Val Ala Arg Gln 2275 2280 2285
- Thr Ser Gln Ile Gly Gly Ser Ser Lys Ala Pro Ser Arg Ser Gly Ser 2290 2295 2300
- Arg Asp Ser Thr Pro Ser Arg Pro Ala Gln Gln Pro Leu Ser Arg Pro 2305 2310 2315 2320
- Ile Gln Ser Pro Gly Arg Asn Ser Ile Ser Pro Gly Arg Asn Gly Ile 2325 2330 2335
- Ser Pro Pro Asn Lys Leu Ser Gln Leu Pro Arg Thr Ser Ser Pro Ser 2340 2345 2350
- Thr Ala Ser Thr Lys Ser Ser Gly Ser Gly Lys Met Ser Tyr Thr Ser 2355 2360 2365
- Pro Gly Arg Gln Met Ser Gln Gln Asn Leu Thr Lys Gln Thr Gly Leu 2370 2375 2380
- Ser Lys Asn Ala Ser Ser Ile Pro Arg Ser Glu Ser Ala Ser Lys Gly 2385 2390 2395 2400
- Leu Asn Gln Met Asn Asn Gly Asn Gly Ala Asn Lys Lys Val Glu Leu 2405 2410 2415
- Ser Arg Met Ser Ser Thr Lys Ser Ser Gly Ser Glu Ser Asp Arg Ser 2420 2425 2430
- Glu Arg Pro Val Leu Val Arg Gln Ser Thr Phe Ile Lys Glu Ala Pro 2435 2440 2445
- Ser Pro Thr Leu Arg Arg Lys Leu Glu Glu Ser Ala Ser Phe Glu Ser 2450 2455 2460
- Leu Ser Pro Ser Ser Arg Pro Ala Ser Pro Thr Arg Ser Gln Ala Gln 2465 2470 2475 2480
- Thr Pro Val Leu Ser Pro Ser Leu Pro Asp Met Ser Leu Ser Thr His 2485 2490 2495
- Ser Ser Val Gln Ala Gly Gly Trp Arg Lys Leu Pro Pro Asn Leu Ser 2500 2505 2510
- Pro Thr Ile Glu Tyr Asn Asp Gly Arg Pro Ala Lys Arg His Asp Ile 2515 2520 2525
- Ala Arg Ser His Ser Glu Ser Pro Ser Arg Leu Pro Ile Asn Arg Ser 2530 2535 2540
- Gly Thr Trp Lys Arg Glu His Ser Lys His Ser Ser Ser Leu Pro Arg 2545 2550 2555 2560

- Val Ser Thr Trp Arg Arg Thr Gly Ser Ser Ser Ser Ile Leu Ser Ala 2565 2570 2575
- Ser Ser Glu Ser Ser Glu Lys Ala Lys Ser Glu Asp Glu Lys His Val 2580 2585 2590
- Asn Ser Ile Ser Gly Thr Lys Gln Ser Lys Glu Asn Gln Val Ser Ala 2595 2600 2605
- Lys Gly Thr Trp Arg Lys Ile Lys Glu Asn Glu Phe Ser Pro Thr Asn 2610 2615 2620
- Ser Thr Ser Gln Thr Val Ser Ser Gly Ala Thr Asn Gly Ala Glu Ser 2625 2630 2635 2640
- Lys Thr Leu Ile Tyr Gln Met Ala Pro Ala Val Ser Lys Thr Glu Asp 2645 2650 2655
- Val Trp Val Arg Ile Glu Asp Cys Pro Ile Asn Asn Pro Arg Ser Gly 2660 2670
- Arg Ser Pro Thr Gly Asn Thr Pro Pro Val Ile Asp Ser Val Ser Glu 2675 2680 2685
- Lys Ala Asn Pro Asn Ile Lys Asp Ser Lys Asp Asn Gln Ala Lys Gln 2690 2695 2700
- Asn Val Gly Asn Gly Ser Val Pro Met Arg Thr Val Gly Leu Glu Asn 2705 2710 2715 2720
- Arg Leu Thr Ser Phe Ile Gln Val Asp Ala Pro Asp Gln Lys Gly Thr 2725 2730 2735
- Glu Ile Lys Pro Gly Gln Asn Asn Pro Val Pro Val Ser Glu Thr Asn 2740 2745 2750
- Glu Ser Pro Ile Val Glu Arg Thr Pro Phe Ser Ser Ser Ser Ser Ser Ser 2755 2760 2765
- Lys His Ser Ser Pro Ser Gly Thr Val Ala Ala Arg Val Thr Pro Phe 2770 2775 2780
- Asn Tyr Asn Pro Ser Pro Arg Lys Ser Ser Ala Asp Ser Thr Ser Ala 2785 2790 2795 2800
- Arg Pro Ser Gln Ile Pro Thr Pro Val Asn Asn Asn Thr Lys Lys Arg 2805 2810 2815
- Asp Ser Lys Thr Asp Ser Thr Glu Ser Ser Gly Thr Gln Ser Pro Lys 2820 2825 2830
- Arg His Ser Gly Ser Tyr Leu Val Thr Ser Val 2835 2840

(2) INFORMATION FOR SEO ID NO:3: (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 3172 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear (ii) MOLECULE TYPE: cDNA (vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens (vii) IMMEDIATE SOURCE: (B) CLONE: DP1(TB2) (ix) FEATURE: (A) NAME/KEY: CDS (B) LOCATION: 1..630 (xi) SEQUENCE DESCRIPTION: SEO ID NO:3: GCA GTC GCC GCT CCA GTC TAT CCG GCA CTA GGA ACA GCC CCG GGN GGC 48 Ala Val Ala Ala Pro Val Tyr Pro Ala Leu Gly Thr Ala Pro Gly Gly GAG ACG GTC CCC GCC ATG TCT GCG GCC ATG AGG GAG AGG TTC GAC CGG 96 Glu Thr Val Pro Ala Met Ser Ala Ala Met Arg Glu Arg Phe Asp Arg 20 25 TTC CTG CAC GAG AAG AAC TGC ATG ACT GAC CTT CTG GCC AAG CTC GAG Phe Leu His Glu Lys Asn Cys Met Thr Asp Leu Leu Ala Lys Leu Glu 35 GCC AAA ACC GGC GTG AAC AGG AGC TTC ATC GCT CTT GGT GTC ATC GGA 192 Ala Lys Thr Gly Val Asn Arg Ser Phe Ile Ala Leu Gly Val Ile Gly 50 CTG GTG GCC TTG TAC CTG GTG TTC GGT TAT GGA GCC TCT CTC TGC 240 Leu Val Ala Leu Tyr Leu Val Phe Gly Tyr Gly Ala Ser Leu Leu Cys 65 75 AAC CTG ATA GGA TTT GGC TAC CCA GCC TAC ATC TCA ATT AAA GCT ATA 288 Asn Leu Ile Gly Phe Gly Tyr Pro Ala Tyr Ile Ser Ile Lys Ala Ile 90 GAG AGT CCC AAC AAA GAA GAT GAT ACC CAG TGG CTG ACC TAC TGG GTA 336 Glu Ser Pro Asn Lys Glu Asp Asp Thr Gln Trp Leu Thr Tyr Trp Val 105

384

432

GTG TAT GGT GTG TTC AGC ATT GCT GAA TTC TTC TCT GAT ATC TTC CTG

Val Tyr Gly Val Phe Ser Ile Ala Glu Phe Phe Ser Asp Ile Phe Leu 115 120 125

TCA TGG TTC CCC TTC TAC TAC ATG CTG AAG TGT GGC TTC CTG TTG TGG

Ser Trp Phe Pro Phe Tyr Tyr Met Leu Lys Cys Gly Phe Leu Leu Trp 130 135 140	
TGC ATG GCC CCG AGC CCT TCT AAT GGG GCT GAA CTG CTC TAC AAG CGC Cys Met Ala Pro Ser Pro Ser Asn Gly Ala Glu Leu Leu Tyr Lys Arg 145	480
ATC ATC CGT CCT TTC TTC CTG AAG CAC GAG TCC CAG ATG GAC AGT GTG Ile Ile Arg Pro Phe Phe Leu Lys His Glu Ser Gln Met Asp Ser Val 165 170 175	528
GTC AAG GAC CTT AAA GAC AAG TCC AAA GAG ACT GCA GAT GCC ATC ACT Val Lys Asp Leu Lys Asp Lys Ser Lys Glu Thr Ala Asp Ala Ile Thr 180 185 190	576
AAA GAA GCG AAG AAA GCT ACC GTG AAT TTA CTG GGT GAA GAA AAG AAG Lys Glu Ala Lys Lys Ala Thr Val Asn Leu Leu Gly Glu Glu Lys Lys 195 200 205	624
AGC ACC TAAACCAGAC TAAACCAGAC TGGATGGAAA CTTCCTGCCC TCTCTGTACC Ser Thr 210	680
TTCCTACTGG AGCTTGATGT TATATTAGGG ACTGTGGTAT AATTATTTTA ATAATGTTGC	740
CTTGGAAACA TTTTTGAGAT ATTAAAGATT GGAATGTGTT GTAAGTTTCT TTGCTTACTT	800
TTACTGTCTA TATATAGG GAGCACTTTA AACTTAATGC AGTGGGCAGT GTCCACGTTT	860
TTGGAAAATG TATTTTGCCT CTGGGTAGGA AAAGATGTAT GTTGCTATCC TGCAGGAAAT	920
ATAAACTTAA AATAAAATTA TATACCCCAC AGGCTGTGTA CTTTACTGGG CTCTCCCTGC	980
ACGSATTTC TCTGTAGTTA CATTTAGGRT AATCTTTATG GTTCTACTTC CTRTAATGTA	1040
CAATTTATA TAATTCNGRA ATGTTTTTAA TGTATTTGTG CACATGTACA TATGGAAATG	1100
TTACTGTCTG ACTACANCAT GCATCATGCT CATGGGGAGG GAGCAGGGGA AGGTTGTATG	1160
TGTCATTTAT AACTTCTGTA CAGTAAGACC ACCTGCCAAA AGCTGGAGGA ACCATTGTGC	1220
TGGTGTGGTC TACTAAATAA TACTTTAGGA AATACGTGAT TAATATGCAA GTGAACAAAG	1280
TGAGAAATGA AATCGAATGG AGATTGGCCT GGTTGTTTCC GTAGTATATG GCATATGAAT	1340
ACCAGGATAG CTTTATAAAG CAGTTAGTTA GTTAGTTACT CACTCTAGTG ATAAATCGGG	1400
AAATTTACAC ACACACACA ACACACACA ACACACACA	1460
AGTACCCTGT AACTCTCAAT TCCCTGAAAA ACTAGTAATA CTGTCTTATC TGCTATAAAC	1520
TTTACATATT TGTCTATTGT CAAGATGCTA CANTGGAMNC CATTTCTGGT TTTATCTTCA	1580
NAGSGGAGAN ACATGTTGAT TTAGTCTTCT TTCCCAATCT TCTTTTTTAA MCCAGTTTNA	1640

GGMNCTTCTG RAGATTTGYC CACCTCTGAT TACATGTATG TTCTYGTTTG TATCATKAGC 1700 AACAACATGC TAATGRCGAC ACCTAGCTCT RAGMGCAATT CTGGGAGANT GARAGGNWGT 1760 ATARAGTMNC CCATAATCTG CTTGGCAATA GTTAAGTCAA TCTATCTTCA GTTTTTCTCT 1820 GGCCTTTAAG GTCAAACACA AGAGGCTTCC CTAGTTTACA AGTCAGAGTC ACTTGTAGTC 1880 CATTTAAATG CCCTCATCCG TATTCTTTGT GTTGATAAGC TGCACAKGAC TACATAGTAA 1940 GTACAGANCA GTAAAGTTAA NNCGGATGTC TCCATTGATC TGCCAANTCG NTATAGAGAG 2000 CAATTTGTCT GGACTAGAAA ATCTGAGTTT TACACCATAC TGTTAAGAGT CCTTTTGAAT 2060 TAAACTAGAC TAAAACAAGT GTATAACTAA ACTAACAAGA TTAAATATCC AGCCAGTACA 2120 GTATTTTTTA AGGCAAATAA AGATGATTAG CTCACCTTGA GNTAACAATC AGGTAAGATC 2180 ATNACAATGT CTCATGATGT NAANAATATT AAAGATATCA ATACTAAGTG ACAGTATCAC 2240 NNCTAATATA ATATGGATCA GAGCATTTAT TTTGGGGAGG AAAACAGTGG TGATTACCGG 2300 CATTTTATTA AACTTAAAAC TTTGTAGAAA GCAAACAAAA TTGTTCTTGG GAGAAAATCA 2360 ACTTTTAGAT TAAAAAATT TTAAGTAWCT AGGAGTATTT AAATCCTTTT CCCATAAATA 2420 AAAGTACAGT TTTCTTGGTG GCAGAATGAA AATCAGCAAC NTCTAGCATA TAGACTATAT 2480 AATCAGATTG ACAGCATATA GAATATATTA TCAGACAAGA TGAGGAGGTA CAAAAGTTAC 2540 TATTGCTCAT AATGACTTAC AGGCTAAAAN TAGNTNTAAA ATACTATATT AAATTCTGAA 2600 TGCAATTTTT TTTTGTTCCC TTGAGACCAA AATTTAAGTT AACTGTTGCT GGCAGTCTAA 2660 GTGTAAATGT TAACAGCAGG AGAAGTTAAG AATTGAGCAG TTCTGTTGCA TGATTTCCCA 2720 AATGAAATAC TGCCTTGGCT AGAGTTTGAA AAACTAATTG AGCCTGTGCC TGGCTAGAAA 2780 ACAAGCGTTT ATTTGAATGT GAATAGTGTT TCAAAGGTAT GTAGTTACAG AATTCCTACC 2840 AAACAGCTTA AATTCTTCAA GAAAGAATTC CTGCAGCAGT TATTCCCTTA CCTGAAGGCT 2900 TCAATCATTT GGATCAACAA CTGCTACTCT CGGGAAGACT CCTCTACTCA CAGCTGAAGA 2960 AAATGAGCAC ACCCTTCACA CTGTTATCAC CTATCCTGAA GATGTGATAC ACTGAATGGA 3020 AATAAATAGA TGTAAATAAA ATTGAGWTCT CATTTAAAAA AAACCATGTG CCCAATGGGA 3080 AAATGACCTC ATGTTGTGGT TTAAACAGCA ACTGCACCCA CTAGCACAGC CCATTGAGCT 3140 ANCCTATATA TACATCTCTG TCAGTGCCCC TC 3172

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 210 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:
- Ala Val Ala Ala Pro Val Tyr Pro Ala Leu Gly Thr Ala Pro Gly Gly
 1 5 10 15
- Glu Thr Val Pro Ala Met Ser Ala Ala Met Arg Glu Arg Phe Asp Arg
 20 25 30
- Phe Leu His Glu Lys Asn Cys Met Thr Asp Leu Leu Ala Lys Leu Glu 35 40 45
- Ala Lys Thr Gly Val Asn Arg Ser Phe Ile Ala Leu Gly Val Ile Gly 50 55 60
- Leu Val Ala Leu Tyr Leu Val Phe Gly Tyr Gly Ala Ser Leu Leu Cys
 65 70 75. 80
- Asn Leu Ile Gly Phe Gly Tyr Pro Ala Tyr Ile Ser Ile Lys Ala Ile 85 90 95
- Glu Ser Pro Asn Lys Glu Asp Asp Thr Gln Trp Leu Thr Tyr Trp Val
- Val Tyr Gly Val Phe Ser Ile Ala Glu Phe Phe Ser Asp Ile Phe Leu 115 120 125
- Ser Trp Phe Pro Phe Tyr Tyr Met Leu Lys Cys Gly Phe Leu Leu Trp 130 135 140
- Cys Met Ala Pro Ser Pro Ser Asn Gly Ala Glu Leu Leu Tyr Lys Arg 145 150 155 160
- Ile Ile Arg Pro Phe Phe Leu Lys His Glu Ser Gln Met Asp Ser Val 165 170 175
- Val Lys Asp Leu Lys Asp Lys Ser Lys Glu Thr Ala Asp Ala Ile Thr 180 185 190
- Lys Glu Ala Lys Lys Ala Thr Val Asn Leu Leu Gly Glu Glu Lys Lys 195 200 205

Ser Thr 210

- (2) INFORMATION FOR SEQ ID NO:5:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 434 amino acids

- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: TB1
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:
- Val Ala Pro Val Val Val Gly Ser Gly Arg Ala Pro Arg His Pro Ala 1 5 10 15
- Pro Ala Ala Met His Pro Arg Arg Pro Asp Gly Phe Asp Gly Leu Gly 20 25 30
- Tyr Arg Gly Gly Ala Arg Asp Glu Gln Gly Phe Gly Gly Ala Phe Pro 35 40 45
- Ala Arg Ser Phe Ser Thr Gly Ser Asp Leu Gly His Trp Val Thr Thr 50 55 60
- Pro Pro Asp Ile Pro Gly Ser Arg Asn Leu His Trp Gly Glu Lys Ser 65 70 75 80
- Pro Pro Tyr Gly Val Pro Thr Thr Ser Thr Pro Tyr Glu Gly Pro Thr 85 90 95
- Glu Glu Pro Phe Ser Ser Gly Gly Gly Gly Ser Val Gln Gly Gln Ser 100 105 110
- Ser Glu Gln Leu Asn Arg Phe Ala Gly Phe Gly Ile Gly Leu Ala Ser 115 120 125
- Leu Phe Thr Glu Asn Val Leu Ala His Pro Cys Ile Val Leu Arg Arg 130 135 140
- Gln Cys Gln Val Asn Tyr His Ala Gln His Tyr His Leu Thr Pro Phe 145 150 155 160
- Thr Val Ile Asn Ile Met Tyr Ser Phe Asn Lys Thr Gln Gly Pro Arg 165 170 175
- Ala Leu Trp Lys Gly Met Gly Ser Thr Phe Ile Val Gln Gly Val Thr 180 185 190
- Leu Gly Ala Glu Gly Ile Ile Ser Glu Phe Thr Pro Leu Pro Arg Glu 195 200 205

Val Leu His Lys Trp Ser Pro Lys Gln Ile Gly Glu His Leu Leu Leu 210 215 220

Lys Ser Leu Thr Tyr Val Val Ala Met Pro Phe Tyr Ser Ala Ser Leu 225 230 235 240

Ile Glu Thr Val Gln Ser Glu Ile Ile Arg Asp Asn Thr Gly Ile Leu 245 250 255

Glu Cys Val Lys Glu Gly Ile Gly Arg Val Ile Gly Met Gly Val Pro 260 265 270

His Ser Lys Arg Leu Leu Pro Leu Leu Ser Leu Ile Phe Pro Thr Val 275 280 285

Leu His Gly Val Leu His Tyr Ile Ile Ser Ser Val Ile Gln Lys Phe 290 295 300

Val Leu Leu Ile Leu Lys Arg Lys Thr Tyr Asn Ser His Leu Ala Glu 305 310 315 320

Ser Thr Ser Pro Val Gln Ser Met Leu Asp Ala Tyr Phe Pro Glu Leu 325 330 335

Ile Ala Asn Phe Ala Ala Ser Leu Cys Ser Asp Val Ile Leu Tyr Pro 340 345 350

Leu Glu Thr Val Leu His Arg Leu His Ile Gln Gly Thr Arg Thr Ile 355 360 365

Ile Asp Asn Thr Asp Leu Gly Tyr Glu Val Leu Pro Ile Asn Thr Gln 370 375 380

Tyr Glu Gly Met Arg Asp Cys Ile Asn Thr Ile Arg Gln Glu Gly 385 390 395 400

Val Phe Gly Phe Tyr Lys Gly Phe Gly Ala Val Ile Ile Gln Tyr Thr 405 410 415

Leu His Ala Ala Val Leu Gln Ile Thr Lys Ile Ile Tyr Ser Thr Leu 420 425 430

Leu Gln

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 185 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: YS-39(TB2)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:
- Glu Leu Arg Arg Phe Asp Arg Phe Leu His Glu Lys Asn Cys Met Thr 1 5 10 15
- Asp Leu Leu Ala Lys Leu Glu Ala Lys Thr Gly Val Asn Arg Ser Phe 20 25 30
- Ile Ala Leu Gly Val Ile Gly Leu Val Ala Leu Tyr Leu Val Phe Gly 35 40 45
- Tyr Gly Ala Ser Leu Leu Cys Asn Leu Ile Gly Phe Gly Tyr Pro Ala 50 55 60
- Tyr Ile Ser Ile Lys Ala Ile Glu Ser Pro Asn Lys Glu Asp Asp Thr 65 70 75 80
- Gln Trp Leu Thr Tyr Trp Val Val Tyr Gly Val Phe Ser Ile Ala Glu 85 90 95
- Phe Phe Ser Asp Ile Phe Leu Ser Trp Phe Pro Phe Tyr Tyr Ile Leu 100 105 110
- Lys Cys Gly Phe Leu Leu Trp Cys Met Ala Pro Ser Pro Ser Asn Gly
 115 120 125
- Ala Glu Leu Leu Tyr Lys Arg Ile Ile Arg Pro Phe Phe Leu Lys His 130 135 140
- Glu Ser Gln Met Asp Ser Val Val Lys Asp Leu Lys Asp Lys Ala Lys 145 150 155 160
- Glu Thr Ala Asp Ala Ile Thr Lys Glu Ala Lys Lys Ala Thr Val Asn 165 170 175

Leu Leu Gly Glu Glu Lys Lys Ser Thr 180

- (2) INFORMATION FOR SEQ ID NO:7:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2843 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein

- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: APC
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:
- Met Ala Ala Ala Ser Tyr Asp Gln Leu Leu Lys Gln Val Glu Ala Leu 1 5 10 15
- Lys Met Glu Asn Ser Asn Leu Arg Gln Glu Leu Glu Asp Asn Ser Asn 20 25 30
- His Leu Thr Lys Leu Glu Thr Glu Ala Ser Asn Met Lys Glu Val Leu 35 40 45
- Lys Gln Leu Gln Gly Ser Ile Glu Asp Glu Ala Met Ala Ser Ser Gly 50 55 60
- Gln Ile Asp Leu Leu Glu Arg Leu Lys Glu Leu Asn Leu Asp Ser Ser 65 70 75 80
- Asn Phe Pro Gly Val Lys Leu Arg Ser Lys Met Ser Leu Arg Ser Tyr 85 90 95
- Gly Ser Arg Glu Gly Ser Val Ser Ser Arg Ser Gly Glu Cys Ser Pro 100 105 110
- Val Pro Met Gly Ser Phe Pro Arg Arg Gly Phe Val Asn Gly Ser Arg
 115 120 125
- Glu Ser Thr Gly Tyr Leu Glu Glu Leu Glu Lys Glu Arg Ser Leu Leu 130 135 140
- Leu Ala Asp Leu Asp Lys Glu Glu Lys Glu Lys Asp Trp Tyr Tyr Ala 145 150 155 160
- Gln Leu Gln Asn Leu Thr Lys Arg Ile Asp Ser Leu Pro Leu Thr Glu 165 170 175
- Asn Phe Ser Leu Gln Thr Asp Met Thr Arg Arg Gln Leu Glu Tyr Glu 180 185 190
- Ala Arg Gln Ile Arg Val Ala Met Glu Glu Gln Leu Gly Thr Cys Gln
 195 200 205
- Asp Met Glu Lys Arg Ala Gln Arg Arg Ile Ala Arg Ile Gln Gln Ile 210 215 220
- Glu Lys Asp Ile Leu Arg Ile Arg Gln Leu Leu Gln Ser Gln Ala Thr 225 230 235 240

Glu Ala Glu Arg Ser Ser Gln Asn Lys His Glu Thr Gly Ser His Asp 245 250 255

Ala Glu Arg Gln Asn Glu Gly Gln Gly Val Gly Glu Ile Asn Met Ala 260 265 270

Thr Ser Gly Asn Gly Gln Gly Ser Thr Thr Arg Met Asp His Glu Thr 275 280 285

Ala Ser Val Leu Ser Ser Ser Ser Thr His Ser Ala Pro Arg Arg Leu 290 295 300

Thr Ser His Leu Gly Thr Lys Val Glu Met Val Tyr Ser Leu Leu Ser 305 310 315 320

Met Leu Gly Thr His Asp Lys Asp Asp Met Ser Arg Thr Leu Leu Ala 325 330 335

Met Ser Ser Gln Asp Ser Cys Ile Ser Met Arg Gln Ser Gly Cys 340 345 350

Leu Pro Leu Leu Ile Gln Leu Leu His Gly Asn Asp Lys Asp Ser Val 355 360 365

Leu Leu Gly Asn Ser Arg Gly Ser Lys Glu Ala Arg Ala Arg Ala Ser 370 375 380

Ala Ala Leu His Asn Ile Ile His Ser Gln Pro Asp Asp Lys Arg Gly 385 390 395 400

Arg Arg Glu Ile Arg Val Leu His Leu Leu Glu Gln Ile Arg Ala Tyr 405 410 415

Cys Glu Thr Cys Trp Glu Trp Gln Glu Ala His Glu Pro Gly Met Asp 420 425 430

Gln Asp Lys Asn Pro Met Pro Ala Pro Val Glu His Gln Ile Cys Pro 435 440 445

Ala Val Cys Val Leu Met Lys Leu Ser Phe Asp Glu Glu His Arg His 450 455 460

Ala Met Asn Glu Leu Gly Gly Leu Gln Ala Ile Ala Glu Leu Leu Gln 465 470 475 480

Val Asp Cys Glu Met Tyr Gly Leu Thr Asn Asp His Tyr Ser Ile Thr
485 490 495

Leu Arg Arg Tyr Ala Gly Met Ala Leu Thr Asn Leu Thr Phe Gly Asp 500 505 510

Val Ala Asn Lys Ala Thr Leu Cys Ser Met Lys Gly Cys Met Arg Ala 515 520 525

Leu Val Ala Gln Leu Lys Ser Glu Ser Glu Asp Leu Gln Gln Val Ile

530 535 540

Ala 545	Ser	Val	Leu	Arg	Asn 550	Leu	Ser	Trp	Arg	Ala 555	Asp	Val	Asn	Ser	_
	Thr	Leu	Arg			Gly	Ser	Val	Lys		Leu	Met	Glu	Cys	560 Ala
Leu	Glu	Val	Lys	565 Lys	Glu	Ser	Thr	Leu	570 Lvs	Ser	Val	Leu	Ser	575 Ala	T.eu
			580					585					590		
Trp	Asn	ьеи 595	Ser	Ala	His	Cys	Thr 600	Glu	Asn	Lys	Ala	Asp 605	Ile	Cys	Ala
Val	Asp 610	Gly	Ala	Leu	Ala	Phe 615	Leu	Val	Gly	Thr	Leu 620	Thr	Tyr	Arg	Ser
Gln 625	Thr	Asn	Thr	Leu	Ala 630	Ile	Ile	Glu	Ser	Gly 635	Gly	Gly	Ile	Leu	Arg 640
Asn	Val	Ser	Ser	Leu 645	Ile	Ala	Thr	Asn	Glu 650	Asp	His	Arg	Gln	Ile 655	Leu
Arg	Glu	Asn	Asn 660	Cys	Leu	Gln	Thr	Leu 665	Leu	Gln	His	Leu	Lys 670	Ser	His

Ser Leu Thr Ile Val Ser Asn Ala Cys Gly Thr Leu Trp Asn Leu Ser 675 680 685

Ala Arg Asn Pro Lys Asp Gln Glu Ala Leu Trp Asp Met Gly Ala Val 690 695 700

Ser Met Leu Lys Asn Leu Ile His Ser Lys His Lys Met Ile Ala Met 705 710 715 720

Gly Ser Ala Ala Leu Arg Asn Leu Met Ala Asn Arg Pro Ala Lys
725 730 735

Tyr Lys Asp Ala Asn Ile Met Ser Pro Gly Ser Ser Leu Pro Ser Leu 740 745 750

His Val Arg Lys Gln Lys Ala Leu Glu Ala Glu Leu Asp Ala Gln His 755 760 765

Leu Ser Glu Thr Phe Asp Asn Ile Asp Asn Leu Ser Pro Lys Ala Ser 770 780

His Arg Ser Lys Gln Arg His Lys Gln Ser Leu Tyr Gly Asp Tyr Val 785 790 795 800

Phe Asp Thr Asn Arg His Asp Asp Asn Arg Ser Asp Asn Phe Asn Thr 805 810 815

Gly Asn Met Thr Val Leu Ser Pro Tyr Leu Asn Thr Thr Val Leu Pro

820 825 830

Ser Ser Ser Ser Arg Gly Ser Leu Asp Ser Ser Arg Ser Glu Lys 835 840 845

Asp Arg Ser Leu Glu Arg Glu Arg Gly Ile Gly Leu Gly Asn Tyr His 850 860

Pro Ala Thr Glu Asn Pro Gly Thr Ser Ser Lys Arg Gly Leu Gln Ile 865 870 875 880

Ser Thr Thr Ala Ala Gln Ile Ala Lys Val Met Glu Glu Val Ser Ala 885 890 895

Ile His Thr Ser Gln Glu Asp Arg Ser Ser Gly Ser Thr Thr Glu Leu
900 905 910

His Cys Val Thr Asp Glu Arg Asn Ala Leu Arg Arg Ser Ser Ala Ala 915 920 925

His Thr His Ser Asn Thr Tyr Asn Phe Thr Lys Ser Glu Asn Ser Asn 930 935 940

Arg Thr Cys Ser Met Pro Tyr Ala Lys Leu Glu Tyr Lys Arg Ser Ser 945 950 955 960

Asn Asp Ser Leu Asn Ser Val Ser Ser Ser Asp Gly Tyr Gly Lys Arg 965 970 975

Gly Gln Met Lys Pro Ser Ile Glu Ser Tyr Ser Glu Asp Asp Glu Ser 980 985 990

Lys Phe Cys Ser Tyr Gly Gln Tyr Pro Ala Asp Leu Ala His Lys Ile 995 1000 1005

His Ser Ala Asn His Met Asp Asp Asn Asp Gly Glu Leu Asp Thr Pro 1010 1015 1020

Ile Asn Tyr Ser Leu Lys Tyr Ser Asp Glu Gln Leu Asn Ser Gly Arg
1025 1030 1035 1040

Gln Ser Pro Ser Gln Asn Glu Arg Trp Ala Arg Pro Lys His Ile Ile 1045 1050 1055

Glu Asp Glu Ile Lys Gln Ser Glu Gln Arg Gln Ser Arg Asn Gln Ser 1060 1065 1070

Thr Thr Tyr Pro Val Tyr Thr Glu Ser Thr Asp Asp Lys His Leu Lys 1075 1080 1085

Phe Gln Pro His Phe Gly Gln Gln Glu Cys Val Ser Pro Tyr Arg Ser 1090 · 1095 1100

Arg Gly Ala Asn Gly Ser Glu Thr Asn Arg Val Gly Ser Asn His Gly

1105 1110 1115 1120

Ile Asn Gln Asn Val Ser Gln Ser Leu Cys Gln Glu Asp Asp Tyr Glu
1125 1130 1135

- Asp Asp Lys Pro Thr Asn Tyr Ser Glu Arg Tyr Ser Glu Glu Glu Gln 1140 1145 1150
- His Glu Glu Glu Arg Pro Thr Asn Tyr Ser Ile Lys Tyr Asn Glu 1155 1160 1165
- Glu Lys Arg His Val Asp Gln Pro Ile Asp Tyr Ser Leu Lys Tyr Ala 1170 1175 1180
- Thr Asp Ile Pro Ser Ser Gln Lys Gln Ser Phe Ser Phe Ser Lys Ser 1185 1190 1195 1200
- Ser Ser Gly Gln Ser Ser Lys Thr Glu His Met Ser Ser Ser Glu 1205 1210 1215
- Asn Thr Ser Thr Pro Ser Ser Asn Ala Lys Arg Gln Asn Gln Leu His 1220 1225 1230
- Pro Ser Ser Ala Gln Ser Arg Ser Gly Gln Pro Gln Lys Ala Ala Thr 1235 1240 1245
- Cys Lys Val Ser Ser Ile Asn Gln Glu Thr Ile Gln Thr Tyr Cys Val 1250 1255 1260
- Glu Asp Thr Pro Ile Cys Phe Ser Arg Cys Ser Ser Leu Ser Ser Leu 1265 1270 1275 1280
- Ser Ser Ala Glu Asp Glu Ile Gly Cys Asn Gln Thr Thr Gln Glu Ala 1285 1290 1295
- Asp Ser Ala Asn Thr Leu Gln Ile Ala Glu Ile Lys Glu Lys Ile Gly 1300 1310
- Thr Arg Ser Ala Glu Asp Pro Val Ser Glu Val Pro Ala Val Ser Gln
 1315 1320 1325
- His Pro Arg Thr Lys Ser Ser Arg Leu Gln Gly Ser Ser Leu Ser Ser 1330 1335 1340
- Glu Ser Ala Arg His Lys Ala Val Glu Phe Ser Ser Gly Ala Lys Ser 1345 1350 1355 1360
- Pro Ser Lys Ser Gly Ala Gln Thr Pro Lys Ser Pro Pro Glu His Tyr 1365 1370 1375
- Val Gln Glu Thr Pro Leu Met Phe Ser Arg Cys Thr Ser Val Ser Ser 1380 1385 1390
- Leu Asp Ser Phe Glu Ser Arg Ser Ile Ala Ser Ser Val Gln Ser Glu

1395 1400 1405

Pro Cys Ser Gly Met Val Ser Gly Ile Ile Ser Pro Ser Asp Leu Pro 1410 1415 1420

- Asp Ser Pro Gly Gln Thr Met Pro Pro Ser Arg Ser Lys Thr Pro Pro 1425 1430 1435 1440
- Pro Pro Pro Gln Thr Ala Gln Thr Lys Arg Glu Val Pro Lys Asn Lys
 1445 1450 1455
- Ala Pro Thr Ala Glu Lys Arg Glu Ser Gly Pro Lys Gln Ala Ala Val 1460 1465 1470
- Asn Ala Ala Val Gln Arg Val Gln Val Leu Pro Asp Ala Asp Thr Leu 1475 1480 1485
- Leu His Phe Ala Thr Glu Ser Thr Pro Asp Gly Phe Ser Cys Ser Ser 1490 1495 1500
- Ser Leu Ser Ala Leu Ser Leu Asp Glu Pro Phe Ile Gln Lys Asp Val 1505 1510 1515 1520
- Glu Leu Arg Ile Met Pro Pro Val Gln Glu Asn Asp Asn Gly Asn Glu 1525 1530 1535
- Thr Glu Ser Glu Gln Pro Lys Glu Ser Asn Glu Asn Gln Glu Lys Glu 1540 1550
- Ala Glu Lys Thr Ile Asp Ser Glu Lys Asp Leu Leu Asp Asp Ser Asp 1555 1560 1565
- Asp Asp Ile Glu Ile Leu Glu Cys Ile Ile Ser Ala Met Pro 1570 1575 1580
- Thr Lys Ser Ser Arg Lys Ala Lys Lys Pro Ala Gln Thr Ala Ser Lys 1585 1590 1595 1600
- Leu Pro Pro Pro Val Ala Arg Lys Pro Ser Gln Leu Pro Val Tyr Lys
 1605 1610 1615
- Leu Leu Pro Ser Gln Asn Arg Leu Gln Pro Gln Lys His Val Ser Phe 1620 1630
- Thr Pro Gly Asp Asp Met Pro Arg Val Tyr Cys Val Glu Gly Thr Pro 1635 1640 1645
- Ile Asn Phe Ser Thr Ala Thr Ser Leu Ser Asp Leu Thr Ile Glu Ser 1650 1655 1660
- Pro Pro Asn Glu Leu Ala Ala Gly Glu Gly Val Arg Gly Gly Ala Gln 1665 1670 1675 1680
- Ser Gly Glu Phe Glu Lys Arg Asp Thr Ile Pro Thr Glu Gly Arg Ser

1685 1690 1695

Thr Asp Glu Ala Gln Gly Gly Lys Thr Ser Ser Val Thr Ile Pro Glu 1700 1705 1710

- Leu Asp Asp Asn Lys Ala Glu Glu Gly Asp Ile Leu Ala Glu Cys Ile 1715 1720 1725
- Asn Ser Ala Met Pro Lys Gly Lys Ser His Lys Pro Phe Arg Val Lys 1730 1735 1740
- Lys Ile Met Asp Gln Val Gln Gln Ala Ser Ala Ser Ser Ser Ala Pro 1745 1750 1755 1760
- Asn Lys Asn Gln Leu Asp Gly Lys Lys Lys Pro Thr Ser Pro Val 1765 1770 1775
- Lys Pro Ile Pro Gln Asn Thr Glu Tyr Arg Thr Arg Val Arg Lys Asn 1780 1785 1790
- Ala Asp Ser Lys Asn Asn Leu Asn Ala Glu Arg Val Phe Ser Asp Asn 1795 1800 1805
- Lys Asp Ser Lys Lys Gln Asn Leu Lys Asn Asn Ser Lys Asp Phe Asn 1810 1815 1820
- Asp Lys Leu Pro Asn Asn Glu Asp Arg Val Arg Gly Ser Phe Ala Phe 1825 1830 1835 1840
- Asp Ser Pro His His Tyr Thr Pro Ile Glu Gly Thr Pro Tyr Cys Phe 1845 1850 1855
- Ser Arg Asn Asp Ser Leu Ser Ser Leu Asp Phe Asp Asp Asp Val
- Asp Leu Ser Arg Glu Lys Ala Glu Leu Arg Lys Ala Lys Glu Asn Lys 1875 1880 1885
- Glu Ser Glu Ala Lys Val Thr Ser His Thr Glu Leu Thr Ser Asn Gln 1890 1895 1900
- Gln Ser Ala Asn Lys Thr Gln Ala Ile Ala Lys Gln Pro Ile Asn Arg 1905 1910 1915 1920
- Gly Gln Pro Lys Pro Ile Leu Gln Lys Gln Ser Thr Phe Pro Gln Ser 1925 1930 1935
- Ser Lys Asp Ile Pro Asp Arg Gly Ala Ala Thr Asp Glu Lys Leu Gln 1940 1945 1950
- Asn Phe Ala Ile Glu Asn Thr Pro Val Cys Phe Ser His Asn Ser Ser 1955 1960 1965
- Leu Ser Ser Leu Ser Asp Ile Asp Gln Glu Asn Asn Asn Lys Glu Asn 1970 1975 1980

- Glu Pro Ile Lys Glu Thr Glu Pro Pro Asp Ser Gln Gly Glu Pro Ser 1985 1990 1995 2000
- Lys Pro Gln Ala Ser Gly Tyr Ala Pro Lys Ser Phe His Val Glu Asp 2005 2010 2015
- Thr Pro Val Cys Phe Ser Arg Asn Ser Ser Leu Ser Ser Leu Ser Ile
 2020 2025 2030
- Asp Ser Glu Asp Asp Leu Leu Gln Glu Cys Ile Ser Ser Ala Met Pro 2035 2040 2045
- Lys Lys Lys Pro Ser Arg Leu Lys Gly Asp Asn Glu Lys His Ser 2050 2055 2060
- Pro Arg Asn Met Gly Gly Ile Leu Gly Glu Asp Leu Thr Leu Asp Leu 2065 2070 2075 2080
- Lys Asp Ile Gln Arg Pro Asp Ser Glu His Gly Leu Ser Pro Asp Ser 2085 2090 2095
- Glu Asn Phe Asp Trp Lys Ala Ile Gln Glu Gly Ala Asn Ser Ile Val 2100 2105 2110
- Ser Ser Leu His Gln Ala Ala Ala Ala Cys Leu Ser Arg Gln Ala 2115 2120 2125
- Ser Ser Asp Ser Asp Ser Ile Leu Ser Leu Lys Ser Gly Ile Ser Leu 2130 2135 2140
- Gly Ser Pro Phe His Leu Thr Pro Asp Gln Glu Glu Lys Pro Phe Thr 2145 2150 2155 2160
- Ser Asn Lys Gly Pro Arg Ile Leu Lys Pro Gly Glu Lys Ser Thr Leu 2165 2170 2175
- Glu Thr Lys Lys Ile Glu Ser Glu Ser Lys Gly Ile Lys Gly Gly Lys 2180 2185 2190
- Lys Val Tyr Lys Ser Leu Ile Thr Gly Lys Val Arg Ser Asn Ser Glu 2195 2200 2205
- Ile Ser Gly Gln Met Lys Gln Pro Leu Gln Ala Asn Met Pro Ser Ile 2210 2215 2220
- Ser Arg Gly Arg Thr Met Ile His Ile Pro Gly Val Arg Asn Ser Ser 2225 2230 2235 2240
- Ser Ser Thr Ser Pro Val Ser Lys Lys Gly Pro Pro Leu Lys Thr Pro 2245 2250 2255
- Ala Ser Lys Ser Pro Ser Glu Gly Gln Thr Ala Thr Thr Ser Pro Arg 2260 2265 2270

- Gly Ala Lys Pro Ser Val Lys Ser Glu Leu Ser Pro Val Ala Arg Gln 2275 2280 2285
- Thr Ser Gln Ile Gly Gly Ser Ser Lys Ala Pro Ser Arg Ser Gly Ser 2290 2295 2300
- Arg Asp Ser Thr Pro Ser Arg Pro Ala Gln Gln Pro Leu Ser Arg Pro 2305 2310 2315 2320
- Ile Gln Ser Pro Gly Arg Asn Ser Ile Ser Pro Gly Arg Asn Gly Ile 2325 2330 2335
- Ser Pro Pro Asn Lys Leu Ser Gln Leu Pro Arg Thr Ser Ser Pro Ser 2340 2345 2350
- Thr Ala Ser Thr Lys Ser Ser Gly Ser Gly Lys Met Ser Tyr Thr Ser 2355 2360 2365
- Pro Gly Arg Gln Met Ser Gln Gln Asn Leu Thr Lys Gln Thr Gly Leu 2370 2375 2380
- Ser Lys Asn Ala Ser Ser Ile Pro Arg Ser Glu Ser Ala Ser Lys Gly 2385 2390 2395 2400
- Leu Asn Gln Met Asn Asn Gly Asn Gly Ala Asn Lys Lys Val Glu Leu 2405 2410 2415
- Ser Arg Met Ser Ser Thr Lys Ser Ser Gly Ser Glu Ser Asp Arg Ser 2420 2425 2430
- Glu Arg Pro Val Leu Val Arg Gln Ser Thr Phe Ile Lys Glu Ala Pro 2435 2440 2445
- Ser Pro Thr Leu Arg Arg Lys Leu Glu Glu Ser Ala Ser Phe Glu Ser 2450 2455 2460
- Leu Ser Pro Ser Ser Arg Pro Ala Ser Pro Thr Arg Ser Gln Ala Gln 2465 2470 2475 2480
- Thr Pro Val Leu Ser Pro Ser Leu Pro Asp Met Ser Leu Ser Thr His 2485 2490 2495
- Ser Ser Val Gln Ala Gly Gly Trp Arg Lys Leu Pro Pro Asn Leu Ser 2500 2510
- Pro Thr Ile Glu Tyr Asn Asp Gly Arg Pro Ala Lys Arg His Asp Ile 2515 2520 2525
- Ala Arg Ser His Ser Glu Ser Pro Ser Arg Leu Pro Ile Asn Arg Ser 2530 2535 2540
- Gly Thr Trp Lys Arg Glu His Ser Lys His Ser Ser Ser Leu Pro Arg 2545 2550 2555 2560

- Val Ser Thr Trp Arg Arg Thr Gly Ser Ser Ser Ser Ile Leu Ser Ala 2565 2570 2575
- Ser Ser Glu Ser Ser Glu Lys Ala Lys Ser Glu Asp Glu Lys His Val 2580 2585 2590
- Asn Ser Ile Ser Gly Thr Lys Gln Ser Lys Glu Asn Gln Val Ser Ala 2595 2600 2605
- Lys Gly Thr Trp Arg Lys Ile Lys Glu Asn Glu Phe Ser Pro Thr Asn 2610 2615 2620
- Ser Thr Ser Gln Thr Val Ser Ser Gly Ala Thr Asn Gly Ala Glu Ser 2625 2630 2635 2640
- Lys Thr Leu Ile Tyr Gln Met Ala Pro Ala Val Ser Lys Thr Glu Asp 2645 2650 2655
- Val Trp Val Arg Ile Glu Asp Cys Pro Ile Asn Asn Pro Arg Ser Gly 2660 2665 2670
- Arg Ser Pro Thr Gly Asn Thr Pro Pro Val Ile Asp Ser Val Ser Glu 2675 2680 2685
- Lys Ala Asn Pro Asn Ile Lys Asp Ser Lys Asp Asn Gln Ala Lys Gln 2690 2695 2700
- Asn Val Gly Asn Gly Ser Val Pro Met Arg Thr Val Gly Leu Glu Asn 2705 2710 2715 2720
- Arg Leu Asn Ser Phe Ile Gln Val Asp Ala Pro Asp Gln Lys Gly Thr 2725 2730 2735
- Glu Ile Lys Pro Gly Gln Asn Asn Pro Val Pro Val Ser Glu Thr Asn 2740 2745 2750
- Lys His Ser Ser Pro Ser Gly Thr Val Ala Ala Arg Val Thr Pro Phe 2770 2775 2780
- Asn Tyr Asn Pro Ser Pro Arg Lys Ser Ser Ala Asp Ser Thr Ser Ala 2785 2790 2795 2800
- Arg Pro Ser Gln Ile Pro Thr Pro Val Asn Asn Asn Thr Lys Lys Arg 2805 2810 2815
- Asp Ser Lys Thr Asp Ser Thr Glu Ser Ser Gly Thr Gln Ser Pro Lys 2820 2825 2830
- Arg His Ser Gly Ser Tyr Leu Val Thr Ser Val 2835 2840

- (2) INFORMATION FOR SEQ ID NO:8:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: peptide
 - (vii) IMMEDIATE SOURCE:
 - (B) CLONE: ral2(yeast)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Leu Thr Gly Ala Lys Gly Leu Gln Leu Arg Ala Leu Arg Arg Ile Ala 1 5 10 15

Arg Ile Glu Gln Gly Gly Thr Ala Ile Ser Pro Thr Ser Pro Leu
20 25 30

- (2) INFORMATION FOR SEQ ID NO:9:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 29 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: peptide
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
 - (vii) IMMEDIATE SOURCE:
 - (B) CLONE: m3 (mAChR)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Leu Tyr Trp Arg Ile Tyr Lys Glu Thr Glu Lys Arg Thr Lys Glu Leu 1 5 10 15

Ala Gly Leu Gln Ala Ser Gly Thr Glu Ala Glu Thr Glu
20 25

- (2) INFORMATION FOR SEQ ID NO:10:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 29 amino acids
 - (C) STRANDEDNESS: single
 - (B) TYPE: amino acid

- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (vi) ORIGINAL SOURCE:

(A) ORGANISM: Homo sapiens

- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: MCC
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Leu Tyr Pro Asn Leu Ala Glu Glu Arg Ser Arg Trp Glu Lys Glu Leu 1 5 10 15

Ala Gly Leu Arg Glu Glu Asn Glu Ser Leu Thr Ala Met 20 25

- (2) INFORMATION FOR SEQ ID NO:11:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 40 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GTATCAAGAC TGTGACTTTT AATTGTAGTT TATCCATTTT

- (2) INFORMATION FOR SEQ ID NO:12:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 40 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TTTAGAATTT CATGTTAATA TATTGTGTTC TTTTTAACAG

(i) SEQUENCE CHARACTERISTICS:

(2) INFORMATION FOR SEQ ID NO:13:

3

40

(2)	INFO	RMATION FOR SEQ ID NO:16:	
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 40 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: cDNA	
	(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:16:	
NNNI	NNNN.	NN NNNGTCCCTT TTTTTAAAAA AAAAAAATAG	40
(2)	INFO	RMATION FOR SEQ ID NO:17:	
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 40 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: cDNA	
	(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:17:	
GTA	\GTAA	CT TGGCAGTACA ACTTATTGA AACTTTAATA	40
(2)	INFO	RMATION FOR SEQ ID NO:18:	
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 40 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: cDNA	
	(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:18:	
ATAC	CAAGA'	TA TTGATACTTT TTTATTATTT GTGGTTTTAG	40
(2)	TMDO	DMATTON FOR CEO ID NO.10.	

(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 40 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: cDNA	
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:19:	
GTAAGTTA	CT TGTTTCTAAG TGATAAAACA GYGAAGAGCT	40
(2) INFO	RMATION FOR SEQ ID NO:20:	
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 40 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: cDNA	
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:20:	
AAAAATAA	CA TAACTAATTA GGTTTCTTGT TTTATTTTAG	40
(2) INFO	RMATION FOR SEQ ID NO:21:	
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 40 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: cDNA	
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:21:	
GTTAGTAA	AT TSCCTTTTTT GTTTGTGGGT ATAAAAATAG	40
(2) INFO	RMATION FOR SEQ ID NO:22:	
(i)	SEQUENCE CHARACTERISTICS:	

(A) LENGTH: 40 base pairs

(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:	
ACCATTTTG CATGTACTGA TGTTAACTCC ATCTTAACAG	40
(2) INFORMATION FOR SEQ ID NO:23:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 40 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: cDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:	
GTAAATAAAT TATTTTATCA TATTTTTTAA AATTATTTAA	40
(2) INFORMATION FOR SEQ ID NO:24:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 64 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:	
CATGATGTTA TCTGTATTTA CCTATAGTCT AAATTATACC ATCTATAATG TGCTTAATTT	60
TTAG	64
(2) INFORMATION FOR SEQ ID NO:25:	
(i) SEQUENCE CHARACTERISTICS:	

	(A) LENGTH: 52 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
(i:	i) MOLECULE TYPE: cDNA	
(v:	i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(x:	i) SEQUENCE DESCRIPTION: SEQ ID NO:25:	
GTAACA	GAAG ATTACAAACC CTGGTCACTA ATGCCATGAC TACTTTGCTA AG	52
(2) IN	FORMATION FOR SEQ ID NO:26:	
(:	i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 46 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(i:	i) MOLECULE TYPE: cDNA	
(v:	i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
	i) SEQUENCE DESCRIPTION: SEQ ID NO:26:	
	TAAA GTCGTAATTT TGTTTCTAAA CTCATTTGGC CCACAG	46
	FORMATION FOR SEQ ID NO:27: i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 40 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii	i) MOLECULE TYPE: cDNA	
(v:	i) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(x:	i) SEQUENCE DESCRIPTION: SEQ ID NO:27:	
GTATGT	TCTC TATAGTGTAC ATCGTAGTGC ATGTTTCAAA	40
(2) INE	FORMATION FOR SEQ ID NO:28:	
()	i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 56 base pairs	

	(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
	<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</pre>	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:	
CATO	CATTGCT CTTCAAATAA CAAAGCATTA TGGTTTATGT TGATTTTATT TTTCAG	56
(2)	INFORMATION FOR SEQ ID NO:29:	
	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 43 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: cDNA	
	(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:	
GTAA	AGACAAA AATGTTTTT AATGACATAG ACAATTACTG GTG	43
(2)	INFORMATION FOR SEQ ID NO:30:	
	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 40 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: cDNA	
	(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:	
TTAG	GATGATT GTCTTTTCC TCTTGCCCTT TTTAAATTAG	40
(2)	INFORMATION FOR SEQ ID NO:31:	
	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 44 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single	

		(D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: cDNA	
	(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:31:	
GTA'	TGTTT	TT ATAACATGTA TTTCTTAAGA TAGCTCAGGT ATGA	44
(2)	INFO	RMATION FOR SEQ ID NO:32:	
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 54 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: cDNA	
	(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:32:	
GCT'	TGGCT	TC AAGTTGNCTT TTTAATĠATC CTCTATTCTG TATTTAATTT ACAG	54
(2)	INFO	RMATION FOR SEQ ID NO:33:	
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 65 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: cDNA	
	(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
*	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:33:	
GTA	CTATT'	FA GAATTTCACC TGTTTTTCTT TTTTCTCTTT TTCTTTGAGG CAGGGTCTCA	60
CTC	rg		65
(2)	INFO	RMATION FOR SEQ ID NO:34:	
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 52 base pairs (B) TYPE: nucleic acid	

	(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: cDNA	
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:34:	
GCAACTAG	TA TGATTTATG TATAAATTAA TCTAAAATTG ATTAATTTCC AG	52
(2) INFC	RMATION FOR SEQ ID NO:35:	
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 42 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: cDNA	
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:35:	
GTACCTTT	GA AAACATTTAG TACTATAATA TGAATTTCAT GT	42
(2) INFO	RMATION FOR SEQ ID NO:36:	
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 40 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: cDNA	
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:36:	
CCAACTCN	AA TTAGATGACC CATATTCAGA AACTTACTAG	40
(2) INFO	RMATION FOR SEQ ID NO:37:	
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 54 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single	

(ii) I	MOLECULE TYPE: cDNA	
(vi) (ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:37:	
GTATATATA	G AGTTTTATAT TACTTTTAAA GTACAGAATT CATACTCTCA AAAA	54
(2) INFOR	MATION FOR SEQ ID NO:38:	
(i) s	SEQUENCE CHARACTERISTICS: (A) LENGTH: 41 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) M	MOLECULE TYPE: cDNA	
(vi) (DRIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
	SEQUENCE DESCRIPTION: SEQ ID NO:38:	
ATTGTGACC".	TAATTTTGTG ATCTCTTGAT TTTTATTTCA G	41
(2) INFORM	MATION FOR SEQ ID NO:39:	
(i) S	SEQUENCE CHARACTERISTICS: (A) LENGTH: 18 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) N	MOLECULE TYPE: cDNA	
(vi) (ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi) S	SEQUENCE DESCRIPTION: SEQ ID NO:39:	
TCCCCGCCTC	G CCGCTCTC	18
(2) INFORM	MATION FOR SEQ ID NO:40:	
(i) S	SEQUENCE CHARACTERISTICS: (A) LENGTH: 18 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single	

(D) TOPOLOGY: linear

	(D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: cDNA	
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:40:	
GCAGCGGC	CGG CTCCCGTG	18
(2) INFO	ORMATION FOR SEQ ID NO:41:	
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: cDNA	
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:41:	
GTGAACGG	CT CTCATGCTGC	20
(2) INFO	RMATION FOR SEQ ID NO:42:	
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 19 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: cDNA	
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:42:	
ACGTGCGG	GG AGGAATGGA	19
(2) INFO	RMATION FOR SEQ ID NO:43:	
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	

•

	(11)	MOLECULE TYPE: CDNA	
	(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:43:	
ATG	ATATC	TT ACCAAATGAT ATAC	24
(2)	INFO	RMATION FOR SEQ ID NO:44:	
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 23 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: cDNA	
	(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:44:	
TTA	TTCCT.	AC TTCTTCTATA CAG	23
(2)	INFO	RMATION FOR SEQ ID NO:45:	
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: cDNA	
	(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
	(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:45:	
TAC	CCATG	CT GGCTCTTTT C	21
(2)	INFO	RMATION FOR SEQ ID NO:46:	
	(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii)	MOLECULE TYPE: cDNA	

	(V1) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:	
TGG	GGCCATC TTGTTCCTGA	20
(2)	INFORMATION FOR SEQ ID NO:47:	
	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: cDNA	
	<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</pre>	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:	
ACA:	TTAGGCA CAAAGCTTGC AA	22
(2)	INFORMATION FOR SEQ ID NO:48:	
	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: cDNA	
	(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:	
ATC	AAGCTCC AGTAAGAAGG TA	22
(2)	INFORMATION FOR SEQ ID NO:49:	
	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 19 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: cDNA	
	(vi) ORIGINAL SOURCE:	

	(A) ORGANISM: Homo sapiens	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:49:	
TGCGGCTCC	T GGGTTGTTG	19
(2) INFOR	MATION FOR SEQ ID NO:50:	
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: cDNA	
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:50:	
GCCCCTTCC	T TTCTGAGGAC	20
(2) INFOR	MATION FOR SEQ ID NO:51:	
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) I	MOLECULE TYPE: cDNA	
(vi) (ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:51:	
TTTTCTCCT	G CCTCTTACTG C	21
(2) INFOR	MATION FOR SEQ ID NO:52:	
(i) s	SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) P	MOLECULE TYPE: cDNA	
(vi) (ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	

(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:52:	
ATGACACC	CC CCATTCCCTC	20
(2) INFO	RMATION FOR SEQ ID NO:53:	
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: cDNA	
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:53:	
CCACTTAA	AG CACATATATT TAGT	24
(2) INFO	RMATION FOR SEQ ID NO:54:	
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: cDNA	
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:54:	
GTATGGAA	AA TAGTGAAGAA CC	22
(2) INFO	RMATION FOR SEQ ID NO:55:	
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: cDNA	
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:	
TTCTTAAGTC CTGTTTTTCT TTTG	24
(2) INFORMATION FOR SEQ ID NO:56:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 23 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: cDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:	
TTTAGAACCT TTTTTGTGTT GTG	23
(2) INFORMATION FOR SEQ ID NO:57:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: cDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:	
CTCAGATTAT ACACTAAGCC TAAC	24
(2) INFORMATION FOR SEQ ID NO:58:	
(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 22 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:	

CATGTCTCTT ACAGTAGTAC CA	22
(2) INFORMATION FOR SEQ ID NO:59:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: cDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:	
AGGTCCAAGG GTAGCCAAGG	20
(2) INFORMATION FOR SEQ ID NO:60:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 27 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: cDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:	
TAAAAATGGA TAAACTACAA TTAAAAG	27
(2) INFORMATION FOR SEQ ID NO:61:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: cDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:	
AAATACAGAA TCATGTCTTG AAGT	24

• 1

(2) INFORMATION FOR SEQ ID NO:62:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 23 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: cDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:	`
ACACCTAAAG ATGACAATTT GAG	23
(2) INFORMATION FOR SEQ ID NO:63:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: cDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:	
TAACTTAGAT AGCAGTAATT TCCC	24
(2) INFORMATION FOR SEQ ID NO:64:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 23 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: cDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:	
ACAATAAACT GGAGTACACA AGG	23
(2) INFORMATION FOR SEQ ID NO:65:	

(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 23 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: cDNA	
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:65:	
ATAGGTCA	TT GCTTCTTGCT GAT	23
(2) INFO	RMATION FOR SEQ ID NO:66:	
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: cDNA	
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:66:	
TGAATTTT	AA TGGATTACCT AGGT	24
(2) INFO	RMATION FOR SEQ ID NO:67:	
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 25 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: cDNA	
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:67:	
CTTTTTTT	GC TTTTACTGAT TAACG	25
(2) INFO	RMATION FOR SEQ ID NO:68:	
(i)	SEQUENCE CHARACTERISTICS:	

	(A) LENGTH: 27 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
	(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:	
TGTA	AATTCAT TTTATTCCTA ATAGCTC	27
(2)	INFORMATION FOR SEQ ID NO:69:	
	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: cDNA	
	<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</pre>	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:	
GGTA	AGCCATA GTATGATTAT TTCT	24
(2)	INFORMATION FOR SEQ ID NO:70:	
	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: cDNA	
	<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</pre>	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:	
CTAC	CCTATTT TTATACCCAC AAAC	24
(2)	INFORMATION FOR SEQ ID NO:71:	
/	~ ***	

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:	
AAGAAAGCCT ACACCATTTT TGC	23
(2) INFORMATION FOR SEQ ID NO:72:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 23 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: cDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:	
GATCATTCTT AGAACCATCT TGC	23
(2) INFORMATION FOR SEQ ID NO:73:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: cDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:	
ACCTATAGTC TAAATTATAC CATC	24
(2) INFORMATION FOR SEQ ID NO:74:	
(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 20 base pairs	

	(C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
	(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:	
GTC	IGGCAT TAGTGACCAG	20
(2)	INFORMATION FOR SEQ ID NO:75:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
	(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:	
AGT	STAATT TTGTTTCTAA ACTC	24
(2)	INFORMATION FOR SEQ ID NO:76:	
	(i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
	(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:	
TGA	GGACTC GGATTTCACG C	21
(2)	INFORMATION FOR SEQ ID NO:77:	
	(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 23 base pairs(B) TYPE: nucleic acid	

(B) TYPE: nucleic acid

(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: cDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:	
TCATTCACTC ACAGCCTGAT GAC	23
(2) INFORMATION FOR SEQ ID NO:78:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: cDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:	
GCTTTGAAAC ATGCACTACG AT	22
(2) INFORMATION FOR SEQ ID NO:79:	
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
(ii) MOLECULE TYPE: cDNA	
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</pre>	
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:	
AAACATCATT GCTCTTCAAA TAAC	24
(2) INFORMATION FOR SEQ ID NO:80:	
(i) SEQUENCE CHARACTERISTICS:(A) LENGTH: 24 base pairs(B) TYPE: nucleic acid(C) STRANDEDNESS: single	

	(D) TOPOLOGY: linear	
	(ii) MOLECULE TYPE: cDNA	
	(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:	
TAC	CATGATT TAAAAATCCA CCAG	24
(2)	INFORMATION FOR SEQ ID NO:81:	
	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 23 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: cDNA	
	(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:	
GAT	GATTGTC TTTTTCCTCT TGC	23
(2)	INFORMATION FOR SEQ ID NO:82:	
	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	
	(ii) MOLECULE TYPE: cDNA	
	(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
	(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:	
CTG	AGCTATC TTAAGAAATA CATG	24
(2)	INFORMATION FOR SEQ ID NO:83:	
	 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 25 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 	

(11)	MOLECULE TYPE: cDNA	
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:83:	
TTTTAAATO	GA TCCTCTATTC TGTAT	25
(2) INFO	RMATION FOR SEQ ID NO:84:	
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: cDNA	
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:84:	
ACAGAGTCA	AG ACCCTGCCTC AAAG	24
(2) INFOR	RMATION FOR SEQ ID NO:85:	
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 23 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: cDNA	
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:85:	
TTTCTATTC	T TACTGCTAGC ATT	23
(2) INFOR	MATION FOR SEQ ID NO:86:	
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii)	MOLECULE TYPE: cDNA	

(vi) C	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi) S	SEQUENCE DESCRIPTION: SEQ ID NO:86:	
ATACACAGGT	T AAGAAATTAG GA	22
(2) INFORM	MATION FOR SEQ ID NO:87:	
(i) S	SEQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) M	MOLECULE TYPE: cDNA	
(vi) C	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi) S	SEQUENCE DESCRIPTION: SEQ ID NO:87:	
TAGATGACCC	C ATATTCTGTT TC	22
(2) INFORM	(2) INFORMATION FOR SEQ ID NO:88:	
(i) S	EQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) M	OLECULE TYPE: cDNA	
(vi) O	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens	
(xi) S	EQUENCE DESCRIPTION: SEQ ID NO:88:	
CAATTAGGTC	TTTTTGAGAG TA	22
(2) INFORM	ATION FOR SEQ ID NO:89:	
	EQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear	
(ii) M	OLECULE TYPE: cDNA	
(vi) 0	RIGINAL SOURCE:	

(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:89:				
GTTACTGC	GTTACTGCAT ACACATTGTG AC				
(2) INFORMATION FOR SEQ ID NO:90:					
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 23 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear				
(ii)	MOLECULE TYPE: cDNA				
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens				
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:90:				
GCTTTTTG	TT TCCTAACATG AAG	23			
(2) INFO	RMATION FOR SEQ ID NO:91:				
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear				
(ii)	MOLECULE TYPE: cDNA				
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens				
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:91:				
TCTCCCAC	AG GTAATACTCC C	21			
(2) INFO	RMATION FOR SEQ ID NO:92:				
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 21 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear				
(ii)	MOLECULE TYPE: cDNA				
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens				

(A) ORGANISM: Homo sapiens

(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:92:				
GCTAGAACTG AATGGGGTAC G 21					
(2) INFO	(2) INFORMATION FOR SEQ ID NO:93:				
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 22 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear				
(ii)	MOLECULE TYPE: cDNA				
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens				
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:93:				
CAGGACAA	AAA TAATCCTGTC CC	22			
(2) INFO	RMATION FOR SEQ ID NO:94:				
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear				
(ii)	MOLECULE TYPE: cDNA				
(vi)	ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens				
(xi)	SEQUENCE DESCRIPTION: SEQ ID NO:94:				
ATTTTCTT	AG TTTCATTCTT CCTC	24			
(2) INFO	RMATION FOR SEQ ID NO: 95:				
(i)	SEQUENCE CHARACTERISTICS: (A) LENGTH: 25 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear				
(ii)	MOLECULE TYPE: cDNA				
(vi)	ORIGINAL SOURCE:				

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:				
AGAAGGATCC CTTGTGCAGT GTGGA 25				
(2) INFORMATION FOR SEQ ID NO: 96:				
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 24 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 				
(ii) MOLECULE TYPE: cDNA				
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</pre>				
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:				
GACAGGATCC TGAAGCTGAG TTTG	24			
(2) INFORMATION FOR SEQ ID NO: 97:				
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 18 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 				
(ii) MOLECULE TYPE: cDNA				
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</pre>				
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:				
TCAGAAAGTG CTGAAGAG				
(2) INFORMATION FOR SEQ ID NO: 98:				
 (i) SEQUENCE CHARACTERISTICS: (A) LENGTH: 19 base pairs (B) TYPE: nucleic acid (C) STRANDEDNESS: single (D) TOPOLOGY: linear 				
(ii) MOLECULE TYPE: cDNA				
<pre>(vi) ORIGINAL SOURCE: (A) ORGANISM: Homo sapiens</pre>				
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:				

GTTCCAGCAG TGTCACAG

GGA	ATAATTA GGTO	CTCCAA	19
(2)	INFORMATION	N FOR SEQ ID NO: 99:	
	(A) I (B) T (C) S	NCE CHARACTERISTICS: LENGTH: 21 base pairs TYPE: nucleic acid STRANDEDNESS: single TOPOLOGY: linear	
	(ii) MOLECU	JLE TYPE: cDNA	
	(vi) ORIGIN (A) C	NAL SOURCE: ORGANISM: Homo sapiens	
	(xi) SEQUEN	ICE DESCRIPTION: SEQ ID NO:	99:
GCA	AATCCTA AGAG	AGAACA A	21
(2)	INFORMATION	FOR SEQ ID NO: 100:	
	(A) L (B) T (C) S	ICE CHARACTERISTICS: LENGTH: 19 base pairs LYPE: nucleic acid LYRANDEDNESS: single LYPOLOGY: linear	
	(ii) MOLECU	LE TYPE: cDNA	
	(vi) ORIGIN (A) O	NAL SOURCE: PRGANISM: Homo sapiens	
	(xi) SEQUEN	CE DESCRIPTION: SEQ ID NO:	100:
GATO	GCAAGC TTGA	GCCAG	19
(2)	2) INFORMATION FOR SEQ ID NO: 101:		
	(A) L (B) T (C) S	CE CHARACTERISTICS: ENGTH: 18 base pairs TYPE: nucleic acid TRANDEDNESS: single OPOLOGY: linear	
	(ii) MOLECU	LE TYPE: cDNA	
	(vi) ORIGIN	AL SOURCE: RGANISM: Homo sapiens	
	(xi) SEQUEN	CE DESCRIPTION: SEQ ID NO:	101:

- (2) INFORMATION FOR SEQ ID NO: 102:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Homo sapiens
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:

GGGAGATTTC GCTCCTGA